Keep an Eye on Stereotypes – The Impact of Gender Stereotypes (Toward Humans and Robots) on Language Processing

A thesis submitted in partial fulfillment of the requirements for the degree of

doctor rerum naturalium (Dr. rer. nat.)

Department of Psychology,

Bielefeld University

Jasmin Bernotat

Defense Committee

Prof. Dr. Friederike Eyssel

Prof. Dr. Gerd Bohner

Prof. Dr. Gernot Horstmann

August 2021

Für meine Mutter Ursula Bernotat, deren fester Glaube an meine Fähigkeiten und unermüdlicher Einsatz diese zu entfalten es mir ermöglichten, diesen Weg zu gehen.

Für meinen Freund Johannes Rabe, der mich in jeglicher Hinsicht unterstützt und ermutigt, „auf Kurs“ zu bleiben.

Acknowledgements

First of all, I would like to thank my supervisors Prof. Dr. Friederike Eyssel and Prof. Dr. Gerd Bohner who have each supported me in their own very special way on this “journey”. Prof. Friederike Eyssel always encouraged me to “step into the front row” and to unfold my research skills and interests. Her “you can”-motivation was of inestimable value to do so and to face new challenges that helped me to grow in any regard. She brought me in contact with many inspiring researchers around the globe which was an enriching experience for me – on a personal and on an academic level. Prof. Dr. Gerd Bohner always supported me with valuable feedback and advice and with critical comments and questions that helped me to gain a new perspective on my research and to overcome challenges during the research process. A special thank also goes to Prof. Dr. Gernot Horstmann for chairing the defense committee.

I would also like to thank my colleagues from the *Applied Social Psychology and Gender Research* and *Social Psychology and Experimental Research on Gender* work groups for inestimable advice and support in numerous meetings, discussions, and colloquia. With regard to my writing phase during this very special Corona-situation, I am also more than grateful that I could participate in regularly online calls held by my supervisor Prof. Friederike Eyssel and my colleagues from the lab. Those meetings kept me connected to the team and were a great source of inspiration how to make the best out of this very special time.

Nevertheless, I would never have been able to finish this interdisciplinary and ambitious research project without former members of the *Language and Cognition Group* Dr. Michele Burigo, Eva Nunnemann, and Dr. Julia Kröger who showed me how to make my way through the "jungle" of psycholinguistic data and who also supported and motivated me on a personal level. A special word of appreciation and thank also goes to my dear friend Dr. Eduardo Benitez Sandoval who always inspired me as a researcher with his passion for social robotics and who always took time to talk things over a coffee. I think I can say, I was lucky to gain new friends over all this time. Thank you so much, mates!

In this regard, I would like to thank my *Movement* peers Kathrin Engel and Raphaela Becker for providing me with “food-for-thought” and “food-for-soul” inspirations. The colleagues from the former *CSRA* group who have invited me to join their weekly “thesis-writing group” were another source of inspiration and motivation for me. Thank you for forcing me to formulate and to pursue concrete thesis-related aims every week. The fact that I finally succeeded to finish this thesis undoubtedly shows that our meetings were of great help and success.

Further appreciation goes to all the students and research interns who have done a lot of great work that helped me to realize this project. I was very fortunate to have very ambitious students who shared the interest in this research project with me. In this regard, a special thank goes to Janik Sachse for his great work on the robot drawings. I would also like to thank *Bielefelder Nachwuchsfonds* for acknowledging my research skills by giving me a grant and thus time I needed to write my thesis. Moreover, the interdisciplinarity and the close connection to other researchers at *CITEC* were a great source of motivation and inspiration for me. In particular, *CITEC Grad School* and the *Movement* program for female upcoming researchers supported by *Bielefeld University* gave me the opportunity to meet my mentor Prof. Dr. Ute Schmid and my peers who advised and motivated me to finish my thesis and thus to get the “entrance ticket” to research.

Finally, I had to find my own way to get this thesis done. Nevertheless, I would not have been able to do so if I would not have had people around me who supported me the way they did, everyone in her or his own very special way. In this regard, a special thank goes to my family and friends who have shared joy and pain with me on my “journey” to get this thesis done and who all believed in my skills and encouraged me to unfold them in the best possible way. Especially my mother’s confidence in my skills and her deep faith in things to finally turn out at their best and my partner’s motivation to live the life I want as a person and as a researcher were of inestimable support. Thank you, without you I would not be the person I am today.

Though I addressed words of thank to many people in this section, I still have the impression that words can hardly express the gratefulness I feel that I had the opportunity to have such great people around me during this time. All the challenges on my way to get this thesis done were worth it because they brought me in contact with all these great people. Moreover, I have been in the lucky position to realize projects that I really loved and that helped me to combine and to develop my interests, skills, and passions. This way, I realized what I want to stand for as a person and as a researcher.

“Fahre fort, übe nicht allein die Kunst, sondern dringe auch in ihr Inneres; sie verdient es. Denn nur die Kunst und die Wissenschaft erhöhen den Menschen bis zur Gottheit.”

[Don't only practice your art, but force your way into its secrets; it deserves that. For art and science can raise (wo)man to the Divine.]

(Ludwig van Beethoven, July 17th, 1812)

Contents

Abstract

In the present thesis, a set of visual world eye tracking experiments was presented in which theories from the fields of psycholinguistics, social psychology, and social robotics were combined. Doing so, four main research aims were pursued: First, I investigated whether adverbs' (Experiment 1 and Experiment 3) and main verbs' (Experiment 2 and Experiment 4) gender-stereotypicality would guide participants' visual attention to a target whose gender matched the adverbs' or the main verbs' gender-stereotypicality. Second, I tested whether the effect of gender-stereotypicality would be enhanced by a male (vs. female) speaker voice. Third, the effects of ambivalent sexism, normative gender role orientation, motivation to control for sexist responses, and social desirability on language processing were explored. Fourth, investigations were extended to the field of social robotics (Experiment 3 and Experiment 4).

The adverbs' and the main verbs' gender-stereotypicality and the speaker voice did not guide participants' visual attention to a gender-matching target as hypothesized. Nevertheless, gender-stereotypicality and speaker voice seemed to have affected language processing to some extent. More precisely, in Experiment 1, participants looked at the character whose gender matched the adverbs' gender-stereotypicality when listening to a female speaker voice, while they looked at the stereotype-inconsistent character when listening to a male speaker voice. In Experiment 3, participants tended to show the hypothesized fixation pattern to look at the target robot whose gender matched the adverbs' gender-stereotypicality, particularly when listening to a speaker whose gender matched the adverbs' gender-stereotypicality. Fixation patterns in Experiment 2 and Experiment 4 suggest that participants seemed to have had difficulties to infer the target's gender from the main verbs' gender-stereotypicality. Viewed in conjunction, the results of all four experiments imply that language processing seemed to be driven by two motives: First, participants seemingly attempted to comprehend language content. After having comprehended that the sentences referred to a male or a female target, participants apparently attempted to counter gender stereotypes and sexism which was indicated by looks at the stereotype-inconsistent target. This was striking because in previous literature, it had not been considered that participants' attempts to respond in a non-stereotypical or non-sexist manner could affect language processing. Experiments featuring robot targets furthermore revealed what kind of robots Western participants considered 'typical' robots and how their perceptions of robots per se and of the social category robots represented in terms of gender affected participants' cognitions during language processing. Based on these insights, directions for future research were outlined.

1 General Introduction

Don't think of a pink elephant!

It is commonly known that when being told not to think of a pink elephant, one immediately thinks of a pink elephant (see e.g., Hooper & McHugh, 2013; Wegner et al., 1987). But why is this the case? Although a vast body of research investigated the psychological aspects of thought suppression (e.g., Clark et al., 1991; Hooper & McHugh, 2013; Wegner, 1989; Wegner et al., 1987), cognitive processes during language comprehension can alternatively be referred to, to address this question. To understand language, it is needed to construct a mental representation of what was heard or read (e.g., Barsalou, 1999; Garnham, 1981; Zwaan & Radvansky, 1998). Paradoxically, it is necessary to construct a mental representation of a pink elephant to know what not to think about. To do so, knowledge about the appearance of the color pink and of elephants must be retrieved from memory. Hence, language comprehension not only requires knowledge about grammatical aspects of language. Furthermore, it requires (and is influenced by) people's knowledge about 'the real world', namely, general world-knowledge about the (behavioral) characteristics, appearance, and the physical properties of objects, animals, and people (see Bransford et al., 1972). According to Zwaan and Radvansky (1998), mental representations of verbally described situations can be referred to as mental models (e.g., Garnham, 1981; Johnson-Laird, 1980; 1983) or as situational models (e.g., van Dijk & Kintsch, 1983; Zwaan & Radvanski, 1998). The core of these notions is the same (van Dijk & Kintsch, 1983). Against prevailing views in the early 1980s, both postulated that language comprehension goes beyond constructing and retrieving the mental representation of a text itself (Zwaan & Radvansky, 1998). Johnson-Laird (1980; 1983) viewed mental models as an independent part of mental representations. He described mental models as internal representations of the world, of external objects, facts, and events which enable to understand verbally described phenomena and to make predictions and inferences based on verbal information (Johnson-Laird, 1983, p. 399). These internal representations are constructed and updated incrementally as long as linguistic information is encountered (Garnham, 1981). They are influenced by (and in turn influence) people's knowledge and the way they perceive the world (Johnson-Laird, 1983).

1.1 The 'Birth' of the Visual World Paradigm

Linguistic research has shown that during language processing, people's mental representations are linked to their visual attention (e.g., Altman, 2011; Altman & Kamide, 2007; Cooper, 1974; Huettig & Altman, 2005; 2007; 2011; Huettig & Janse, 2016; Knoeferle & Crocker, 2006; 2007; Tanenhaus et al.,1995). In his 1974 study, Cooper instructed participants to look at drawings of distinct commonplace objects. Simultaneously, he presented spoken prose passages that referred either directly or indirectly to the drawings. Cooper found that participants spontaneously directed their visual attention to those targets that were most closely related to the meaning of the linguistic input they had heard. This shift of visual attention rapidly occurred while an informative word of a prose passage was still pronounced or within 200 ms after its termination. Although Cooper's research seemed promising, it did not gain much attention in the field of psycholinguistics until Tanenhaus and his colleagues (1995) used a similar approach (see Huettig et al.,2011b for a review). Using a head-mounted eye tracker, Tanenhaus et al. (1995) orally instructed participants to move real objects. The instructions contained temporary *syntactic ambiguity* (e.g., ambiguous: "Put the apple on the towel in the box." vs. unambiguous: "Put the apple that is on the towel in the box."). In an ambiguous instruction, the prepositional phrase "on the towel" could be interpreted as a destination where to place the target object or as a modifier which target object to place. The referring visual scene either contained one target object (one-referent visual context) or two target objects (two-referent visual context), while the correct target object was the object to be moved. To illustrate, in a one-referent visual context, an apple was presented on a towel amongst a towel with no object on it, a box, and a distractor object that was unrelated to the verbal instruction. The one- and the two-referent visual context differed therein that in the two-referent visual context, the distractor object was replaced by another apple on a napkin. Thus, when the two-referent visual scene was presented, it was not clear which apple had to be moved until the prepositional phrase "on the towel" specified the correct target object. Thus, unsurprisingly, when an unambiguous instruction referred to a one-referent visual context, participants rarely fixated the incorrect target object. Contrarily, when a two-referent context was presented, both apples were looked at as soon as "the apple" was mentioned in the instruction. This effect was even more pronounced when the instructions contained ambiguous information about the correct target object than when the instructions were unambiguous. Independent of the instructions' syntactic ambiguity, upon listening to "on the towel", participants' visual attention immediately shifted toward the correct target object and to the box where to place it. This indicated that the prepositional phrase "on the towel" was correctly identified as a modifier, not as a destination. It was thus used to resolve ambiguity. These findings imply that participants aimed to establish a reference between verbal and visual information very early during language processing. Doing so, they processed the instructions incrementally and prediction-driven while relevant visual information immediately impacted how verbal input was structured.

1.2 The Visual World Paradigm as a Mirror of Listeners' Mental Representations

Cooper's (1974) and Tanenhaus and colleagues' (1995) experimental methodology to record participants' eye movements while verbal information and a referring visual context are simultaneously presented, is nowadays commonly known as the *visual world paradigm* (e.g., Allopenna et al.,1998; Magnuson et al.,1999; see also Huettig et al.,2011b for a review). Using the visual world paradigm enables to measure participants' visual attention which in turn reveals their cognitive processes. That is, visual attention mirrors the interplay of language perception, memory, and language comprehension while speech is presented (e.g., Cooper, 1974; Just & Carpenter, 1976; Knoeferle & Crocker, 2006; 2007; Tanenhaus et al., 1995; see also Huettig et al.,2011a; Huettig et al.,2011b for reviews). Shifts of visual attention unveil participants' spontaneous associations and anticipations when encountered with spoken language. These changes in gaze patterns over time can be recorded with short latencies (see Barr, 2008; Cooper, 1974; Matin et al.,1993; Tanenhaus et al., 1995). That is, the visual world paradigm is suitable to investigate participants' cognitive processes in 'real-time' without considerable time delays. It thus seems relatively robust against participants' intentional control over their reactions (see Goldberg & Wichansky, 2003; Rayner, 1998). Another advantage of the visual world paradigm is that it does not require participants to perform an unnatural meta-linguistic task (e.g., to decide whether a target stimulus is a noun or a verb) (see e.g., Huettig et al.,2011a; Huettig et al.,2011b). Although, in some visual world experiments participants are asked to move or to click on objects that are displayed on a screen (e.g., Magnuson et al., 1999; Spivey et al.,2002; Tanenhaus et al., 1995). In other cases, the task in a visual world experiment can simply be to pay attention to the visual and the verbal stimuli that are presented (e.g., Altman, 2004; Huettig et al.,2011c; McQueen & Huettig, 2012). Due to these advantages, the visual world paradigm has triggered a vast range of research over the last two decades. In the psycholinguistic field, the visual world paradigm has been used, for instance, to investigate lexical and syntactic ambiguity resolution (e.g., Dahan & Tanenhaus, 2004; Fernandes et al.,2015; Moreno et al.,2015; Spivey et al., 2002; Tanenhaus et al., 1995; Trueswell et al.,1999), language comprehension of abstract vs. concrete words (e.g., Duñabeitia et al.,2009), the impact of physical characteristics of depicted objects on language processing, such as color and shape (e.g., Huettig & Altman, 2011; Sedivy et al.,1999), similarities of presented objects (e.g., Dahan & Tanenhaus, 2005), and the interrelation between a visual context and verbal information and its impact on participants' predictions of semantic information (e.g., Altman & Kamide, 1999; 2007; 2009; Kamide et al.,2003). To illustrate, Altman and Kamide (1999) presented recordings of either "The boy will *move* the cake." or "The boy will *eat* the cake.". While listening to the respective sentence, a visual scene depicting a boy and a cake amongst unrelated distractor objects was shown. The cake was the only edible object that was depicted. When listening to "The boy will *eat* the cake.", participants' visual attention shifted to the depicted cake even before it was mentioned in the sentence. However, upon hearing "The boy will *move* the cake.", participants' visual attention shifted to the depicted cake after it was named in the sentence. These findings demonstrate that verbs can be used as a source of information to guide participants' predictive eye movements to the object the verb most likely refers to. Participants' prediction-driven visual attention in turn mirrors their mental representations of the relation between entities in the real world.

1.3 Language Processing as an Incremental Process: The Interplay of Verbal and Visual Information and Comprehenders' Mental Representations

The presented research characterized language processing as an incremental process which continues as long as verbal information is presented. To build a mental representation of verbally described situations and to predict ongoing input, inferences were drawn from verbal and visual information and comprehenders' (the term is common in linguistics as it addresses readers and listeners equally) existing knowledge. Nonetheless, until Knoeferle and Crocker (2006; 2007; Knoeferle et al.,2014), it was not specified how these sources of information interfered and whether one was preferred over the other.

Using the visual world paradigm, Knoeferle and Crocker (2006, Experiment 2) presented spoken sentences, such as "Den Piloten *verköstigt* vs. *verzaubert* vs. *bespitzelt* gleich der *Detektiv*." ["The pilot *serves-food to* vs. *jinxes* vs. *spies-on* soon the *detective*."]. At the same time, a pilot was depicted amongst a detective and a wizard. The detective who is commonly known as spying on someone, was portrayed as serving food. The wizard who is commonly known as jinxing was portrayed as spying on someone. When listening to "*verköstigt*" [*serves-food to*], participants immediately looked at the detective who was portrayed as serving food. When listening to "*verzaubert*" [*jinxes*], participants looked at the wizard as none of the agents was depicted as jinxing and as it was the only agent that is commonly known as jinxing. That is, when given information unambiguously referred to a target, participants relied on depicted events and on their knowledge about a role's typical actions. But what source of information was preferred when visual information and knowledge competed? Before listening to "*bespitzelt"* [*spies-on*], participants could either rely on their stereotypical knowledge or on the visual scene to predict the sentence's target. After listening to "*bespitzelt"* [*spies-on*], participants' visual attention shifted to the wizard that was depicted as spying the pilot, not to the detective who is commonly known as spying on someone. In all conditions, participants' visual attention shifted to the target agent after it had been named which indicated that they revised their predictions if necessary. Overall, the authors concluded that when both was relevant for language comprehension, participants preferably relied on visual information over stereotypical information. Using a *blank screen paradigm* (see e.g., Altman, 2004; Huettig & Janse, 2016), Knoeferle and Crocker (2007) replicated these findings. They concluded that the preference for visual over stereotypical knowledge depended on the extent to which visual scenes decay from working memory (see Baddeley, 2000; Baddeley & Hitch, 1974; Baddeley et al.,2009 for working memory functions). Based on their findings, Knoeferle and Crocker (2006; 2007) derived the *Coordinated Interplay Account (CIA)* of situated language processing. The CIA describes the interplay of incoming verbal and visual information, comprehenders' mental representations of a verbally described situation stored in working memory, and their general world-knowledge stored in long-term memory. According to the CIA, language is continuously processed in three steps: *sentence interpretation, utterance mediated attention,* and *scene integration.* These steps depend on each other in terms of content and timing. They can occur simultaneously or partly overlapping. The first step – scene interpretation – is initiated as soon as comprehenders encounter the first word of an utterance. Upon listening to (or reading) a word, comprehenders retrieve their linguistic and general world-knowledge, interpret its meaning, verify their interpretation, and predict a following word. These interpretations, verifications, and predictions as well as the referring visual scene that was previously inspected are stored in working memory. They direct comprehenders' visual attention to entities that might be relevant for language comprehension. These shifts of visual attention mark the second step of language comprehension: utterance mediated attention. Finally, the third step – scene integration – proceeds: New relevant visual information is added to the mental representation of the scene stored in working memory which is then revised accordingly. Entities that are no longer depicted then decay. Knoeferle et al. (2014) further emphasized the role of working memory. They acknowledged that comprehenders might differ in their working memory capacity and in processing time. This was one of the first approaches to consider the effects of comprehenders' individual characteristics on language processing which are still mainly disregarded in linguistic research.

1.4 Language Processing as a Social Phenomenon: The Role of Stereotypes and the Social Context

The idea to integrate comprehender characteristics amongst social aspects in linguistic theories was not new (e.g., Labov, 1972; Průcha, 1972; see also De Bernardi, 1994 for an overview). However, these factors were not yet in the research focus until Münster and

Knoeferle (2018) proposed an extension of the CIA – the *Social CIA.* The Social CIA emphasized the impact of comprehender characteristics (e.g., age, educational level, mood, gender stereotypes, beliefs, world-knowledge) and the socially interpreted context (e.g., referents, actions, events, gestures, emotional facial expressions, and speaker voice) on language processing. Their model was derived from an extensive review of psycholinguistic, sociolinguistic, and social psychological literature. With their model, they transferred social psychologic approaches to linguistic research.

In the context of social psychology, particularly the impact of gender stereotypes on language processing was well-researched in visual world experiments (e.g., Guerra et al., 2021; Pyykkönen, et al.; 2010; Rodriguez et al.,2015; 2016) and in reading experiments (e.g., Canal et al.,2015; Carreiras et al.,1996; Duffy & Keir, 2004; Esaulova et al.,2014; Finnegan, et al.; 2015; Gabriel et al.,2017; Garnham et al.,2002; Irmen, 2007; Irmen & Roßberg, 2004; 2006; Wang et al.,2017). According to this line of research, participants need more time to process a sentence when a target's actual gender mismatches participants' knowledge about its gender-stereotypicality. Interpreted in the context of the (Social) CIA, this appears plausible because participants have to revise their initial interpretation of an utterance when subsequent verbal information mismatches their expectations. This phenomenon became known as the *mismatch effect* (e.g., Duffy & Keir, 2004; Kennison & Trofe, 2003).

It also occurred when participants' interpretation of an utterance's social context was inconsistent with their expectations. To form interpretations of an utterance's social context, comprehenders use subtle information about a target's identity. As such, a target's hands were used to infer his or her gender (Rodriguez et al.,2015; 2016) or a speaker's voice served as an indication of gender, age, and socio-economic status (Van Berkum et al.,2008). To illustrate, Rodriguez and her colleagues (2016) showed video sequences in which male (vs. female) hands performed a gender-stereotypical action (e.g., baking a cake, stereotypically female action). A mismatch effect occurred when a target's gender (indicated by her or his hands) was inconsistent with participants' knowledge about an action's gender-stereotypicality (e.g., when male hands were shown as performing a stereotypically female action, such as baking a cake). Similarly, measuring electro-related brain potentials (ERPs), Van Berkum and his colleagues (2008) reported mismatch effects when a sentence's content was inconsistent with participants' expectations about a speaker's gender (e.g., "If only I looked like *Britney* Spears in her latest video." spoken by a male speaker), age (e.g., "Every evening I drink some *wine* before I go to sleep." uttered by a child), and socio-economic status (e.g., "I have a large *tattoo* on my back."

spoken with an upper-class accent). Such mismatches evoked the same brain responses as syntactic anomalies which indicates that on a cognitive level, they equaled grammatical errors (see also Osterhout et al.,1997). More importantly, these experiments demonstrated that comprehenders rapidly form a mental representation of the speaker to predict incoming verbal information. To do so, they use any contextual information available, even if its subtle (see also Giles et al.,1991; Niederhoffer & Pennebaker, 2002). Van Berkum and colleagues (2008) emphasized the relevance of the social context on language interpretation: “The linguistic brain is not just combining words in a context-free semantic universe confined in a single person’s skull. It immediately cares about other people.” (Van Berkum et al., 2008, p. 589).

Although psycholinguists had acknowledged the relevance of social and psychological factors, such as speakers’ and comprehenders’ demographic background, on language processing, language processing is still primarily regarded as a linguistic phenomenon. Van Berkum et al. (2009) criticized that comprehenders’ attitudes, beliefs, moral values, and emotions are still widely neglected in psycholinguistic research. Their criticism was legitimate: Although a vast body of research highlighted the impact of gender stereotypes on language processing, the effects of comprehenders’ gender, sexist attitudes, and their motives remained mainly unconsidered in this context. Social psychologists had acknowledged the impact of attitudes and motives in the context of gender stereotypes (e.g., Glick & Fiske, 1996; 2001; Klonis et al.,2005), but they have neglected that gender stereotypes manifest in language. The psychological factors that underlie language processing and that might explain comprehenders’ perception and attention while encountering gender stereotypes are thus still rarely explored.

1.5 (Gender) Stereotypes and Sexist Attitudes: A Social Psychologic Perspective

In most previous psycholinguistic research, a clear understanding of gender stereotypes and a differentiation to related concepts, such as sexist attitudes, motives, and non-sexist norms, seems to be still missing. To illustrate, because prevailing stereotypes were assumed to be commonly known, they were often equated to general world-knowledge about things and people. For instance, Pyykkönen and her colleagues (2010) referred to gender stereotypes as “one type of knowledge” (p. 126) about roles stereotypically performed by men and women. This equation bears the risk to regard stereotypes as given facts whose correctness is not questioned. Moreover, it neglects that stereotypes are a social phenomenon which is closely related to one’s attitudes and emotions toward a stereotyped group or individual. Therefore, some authors made attempts to differentiate stereotypes from general world-knowledge. To illustrate, Contreras and his colleagues (2012) found that, unlike non-social categories, stereotypes about groups activated brain regions that are involved when people think of social groups’ attributes, form impressions of others, and infer others' beliefs, emotions, opinions, and intentions (see also Molinaro et al.,2016; Norris et al.,2004). This brain activity fits well with a social psychologic definition of stereotypes as beliefs about a social group's attributes (Ashmore & Del Boca, 1979; Eagly & Mladinic, 1989). Stereotypes allow predictions about others' behavior and intentions (i.e., good or bad). This in turn determines the evaluations of and emotions and behavior toward them (Ashmore & Del Boca, 1979; Bodenhausen et al.,2012; Eagly & Mladinic, 1989; Fiske, 2018; Fiske et al.,1999; Fiske et al.,2002).

With regard to gender stereotypes, two fundamental dimensions have been outlined to characterize men and women. These were referred to as *masculinity* and *femininity* (e.g., Bem, 1974; 1981; Berger & Krahé, 2013 furthermore differed between positive and negative attitudes). In more general terms, it was referred to *warmth* and *competence* (*stereotype content model*: Fiske, 2018; Fiske et al.,1999; Fiske et al.,2002; Fiske et al.,2007), *competence* and *morality* (Wojciszke, 1994; Wojciszke et al.,1998), or *agency* and *communion* (Abele, 2003; Abele & Wojciszke, 2007; 2014; Bakan, 1966). Abele and her colleagues (2016) proposed a four-dimensional model. They regarded *competence* and *assertiveness* as two facets of agency and *warmth* and *morality* being facets of communion.

Accordingly, Men are regarded as competent (agentic, Abele, 2003), whereas women are perceived as warm, but incompetent. This perception rather accounts for women in traditional roles (Fiske et al.,1999; Fiske et al.,2002). It has slightly changed over the past decades. Comparing polls on gender-stereotypicality over the last seven decades, Eagly and her colleagues (2019) confirmed that agentic traits were still rather associated with men, while women were still perceived as communal. Though, women's level of both communion and competence increased over time. Their level of competence was even found to be at least as high as men's. These results suggest that a more fine-grained classification of gender-stereotypicality as proposed by Abele and colleagues (2016) might be useful.

Eagly and colleagues (2019) concluded that stereotypes were derived from a group's roles and behavior (see also Koenig & Eagly, 2014; Wood & Eagly, 2012). Particularly female roles have changed drastically over the past decades. Gender stereotypes about how men and women are and what they do have thus changed over time (see also Auster & Ohm, 2000).

Women's roles are closely related to the attitudes and emotions they evoke. Glick and Fiske (1996; 2001) differed between *benevolent sexism* and *hostile sexism.* The former encompasses a subjectively positive view, even an idealization of women and their traditional roles as mother and romantic partner. It is accompanied by paternalistic feelings which elicit seemingly prosocial and protective behavior toward them. The latter encompasses a blatant negative view of women as 'gatekeepers' who aim to master men. Though benevolent and hostile sexism seem to reflect two contrasting views of women, they are two 'sides of the same coin'. Both serve to maintain men's structural power (Glick & Fiske, 1996; 2001). It thus appears logical that women who challenge men's structural power (e.g., by pursuing a career, by being agentic) often experience hostility and discrimination (e.g., Rudman & Glick, 1999; 2001; Rudman et al.,2012; see also Heilman, 2012; Phelan & Rudman, 2010 for reviews).

1.6 The Role of Cultural vs. Personal Beliefs: Counteracting Gender Stereotypes and Prejudices

Gender stereotypes and sexist beliefs are culturally inherited (Devine, 1989; Glick & Fiske, 1996). As such, they are automatically activated as soon as they are encountered (e.g., Chen & Bargh, 1997; Devine, 1989; Zhang et al.,2018). However, culturally shared gender stereotypes and sexist beliefs do not necessarily reflect one's personal beliefs. People were found to differ in the extent to which they consider stereotypes and prejudices as appropriate. When considered inappropriate, people counteracted stereotype content and prejudiced responses under certain conditions: When they were aware of stereotype content, had cognitive resources and time, and when they were motivated to do so (Devine, 1989; Fazio, 2007). People might be motivated to respond non-prejudiced for three reasons: First, because they are internally motivated to do so. That means, they truly consider prejudices as inappropriate and have internalized the aim not to be prejudiced. Second, because they are externally motivated to respond non-prejudiced. That is, they feel pressure to behave non-prejudiced due to social norms. Third, because of both. That is, they consider prejudices as inappropriate and fear disapproval for showing prejudiced behavior (Klonis et al.,2005; see also Plant & Devine, 2009). Thus, plausibly, participants who are internally motivated to respond non-prejudiced were found to control for prejudiced responses independent of others' approval. On the contrary, externally motivated participants controlled for prejudiced responses in the presence of a stereotype target (e.g., Fazio et al.,1995; Lowery et al.,2001) or when in fear others might detect their prejudices (e.g., Klonis et al.,2005; Plant & Devine, 2009). Participants' motivation to respond without prejudice therefore seems to have a crucial impact on the processing of stereotype content. In the context of racial stereotypes, it even turned out a better predictor of counteractive responses than a person's attitudes (Plant & Devine, 2009).

Summing up, the reviewed literature demonstrated that gender stereotypes, sexist attitudes, motives, and the social context (e.g., Fazio et al., 1995; Klonis et al., 2005; Lowery et al., 2001) have a crucial impact on how stereotypes are processed. It is thus essential to consider them in research on language comprehension of gender stereotypes. In turn, because gender stereotypes commonly occur in language, they also need to be considered a linguistic phenomenon.

1.7 Humanlike Robots: Representatives of Social Categories

When investigating social categories, such as gender, prior psycholinguistic research mainly focused on human targets. This approach neglected the fact that gender stereotypes and prejudices can also be transferred to social robots. One might wonder why robots are relevant in this regard. Robots increasingly enter the labor force and private contexts (Müller, 2020). They are considered social agents that will eventually closely interact with humans (see Duffy, 2003; Jackson & Williams, 2019). As such, their design and behavior are modelled after humans. To date, robots at least appear as if they were autonomous and had social skills and the ability to learn. This makes them distinct from other non-human objects and leads people to anthropomorphize them (Breazeal, 2003). That is, to ascribe humanlike traits, emotions, and intentions to them (see Epley et al.,2007).

A robot's human-like appearance and people's tendency to anthropomorphize robots, seem to be a double-edged sword: On the one hand, they enable to predict a robot's behavior which facilitates human-robot interaction (see Duffy, 2003; Epley et al.,2007). On the other hand, a robot's humanlike features, such as facial cues (Eyssel & Hegel, 2012), body shape (Bernotat et al.,2017; 2021; Strait et al.,2017), alleged origin (Eyssel & Kuchenbrandt, 2012), and skin color (Bartneck et al., 2018) led to social categorization in terms of gender and ethnicity. It even resulted in discrimination against robots (see Bartneck et al., 2018; Strait et al., 2017).

To illustrate, in the context of gender stereotypes, Eyssel and Hegel (2012) demonstrated that a male robot, indicated by a short hair part, was perceived as more agentic and deemed more suitable for stereotypically male tasks. Reversely, a female robot was perceived as communal and preferred for stereotypically female tasks. Bernotat, et al. (2017; 2021) varied robot gender by manipulating its waist-to-hip ratio and shoulder width. The authors replicated Eyssel and Hegel's findings only for the female robot. Both robots were deemed equally agentic and suitable for stereotypically female tasks. Bernotat et al. (2017; 2021) exceeded Eyssel and Hegel's research in three aspects: First, they considered the impact of robot gender on participants' trust in human-robot interaction (HRI). By doing so, they were the first who provided a measure of trust in HRI as a two-dimensional concept of *cognitive* and *affective trust*. Cognitive trust denotes one's trust in a robot's functions. Affective trust defines one's trust in a robot's benevolent motives, its care, and concern. The female robot evoked more cognitive and affective trust than its male counterpart. Second, the authors investigated participants' overall perceptions of robots. They found that, independent of the robot's alleged gender, participants indicated higher levels of female robot gender and, in tendency, communal traits than male gender and agentic traits. This was surprising. Because robots were commonly known as male (Jung et al.,2016), one would expect them to be perceived rather as male and agentic. At the same time, participants showed more trust in a robot's functions than in its motives and favored robots for stereotypically male tasks over female tasks. This preference was explained by the finding that stereotypically male tasks (e.g., transporting goods) required less close human-robot interaction than stereotypically female tasks (e.g., caring for elderly). Third, Bernotat et al. (2017; 2021) uncovered that participants' societal beliefs about gender-stereotypical traits, sexist attitudes, gender, attitudes toward robots and technology, and social desirability had affected their evaluations of the robots. Bernotat et al. (2017; 2021) concluded that gendered robots had apparently activated participants' knowledge about and attitudes toward *gender* and *robots* at the same time. That is, robots did apparently not only activate stereotypes associated with the social group they represented, but with the category of robots itself.

1.8 People's View of Robots

To understand how people commonly view robots and what attitudes they share toward them, it needs to be considered that people do not yet have much experience with robots (Bernotat & Eyssel, 2018; Bernotat et al.,2021; de Graaf & Allouch, 2013). In Western Societies, the majority of people knows robots only from the media (Bernotat & Eyssel, 2018; Bernotat et al.,2021; Horstmann & Krämer, 2019; Sandoval et al.,2014). Media portrays robots as intelligent, artificial, humanlike labor forces that support humans in the beginning, until, one day, they turn against them like in Čapek's (1921) early play. With this image in mind, it appears plausible that people hesitate to interact with robots (e.g., Bernotat et al., 2016; Bernotat & Eyssel, 2018; Bernotat et al.,2021; Reich-Stiebert & Eyssel, 2015; 2016) and have ambivalent feelings toward them (e.g., Dang & Lui, 2020; MacDorman et al.,2009; Stapels & Eyssel, 2021). This ambivalence toward robots might be reflected in participants' perceptions of humanlike robots. To illustrate, to some extent, a robot's humanlike appearance might facilitate HRI (Duffy, 2003). However, already in 1970, Mori warned that if a robot exceeds a certain level of humanlikeness, it evokes feelings of eeriness. Likewise, perceptions of a robot's mindfulness (Gray & Wegner, 2012, Stafford et al.,2014) and autonomy (Złotowski et al.,2017) caused feelings of unease (Gray & Wegner, 2012). The results were however inconsistent. Stafford and her colleagues (2014) concluded that high perceived agency and low perceived robot experience caused discomfort. The authors argued that if a robot was perceived as highly agentic, participants feared it could have intentions, make plans, and become independent of their human users. A lack of experience indicated that a robot would not be able to show empathy for human needs and emotions. On the contrary, Gray and Wegner (2012) found that perceived experience resulted in feelings of uncanniness, rather than a humanlike body shape per se. The authors concluded that experience was a fundamentally human characteristic. Taken together, these results confirmed that robots should possess emotionality and sociability only to a certain degree so that human distinctness is not threatened (see also Vanman & Kappas, 2019 for a review). Perceived threat and the need for distinctness also play an important role in human intergroup contact situations (see Mendes et al.,2002; Tamir & Nadler, 2007; see also Hewstone & Greenland, 2000 for a review). Indeed, human-robot interaction might, to some extent, be regarded as an intergroup situation which is coined by people's associations, stereotypes, and attitudes toward robots. Like in human intergroup situations, people's associations and attitudes determine their emotions and behavior in HRI. Therefore, social psychologists and roboticists have called to consider the interplay of robot characteristics (e.g., design, functionality, autonomy), participants' characteristics (e.g., demographics, attitudes, motives, emotions, personality), and the context in which HRI takes place (Bernotat & Eyssel, 2018; Hancock et al., 2011; Hancock et al.,2020; Schaefer, 2013). To illustrate the latter, in most HRI studies, participants articulated resentments toward robots (e.g., Bernotat et al., 2016; Bernotat & Eyssel, 2018; Bernotat et al.,2021; Reich & Eyssel, 2013; Stafford et al., 2014). However, in an anonymous setting, such as while watching videos of robots on *YouTube* (an online platform for videos), people expressed clear discomfort particularly toward highly anthropomorphic (vs. 'prototypical' robotic) robots. People's emotionality and their fear robots might take over were positively correlated. Additionally, alleged female robots received blatant sexist comments (Strait et al.,2017). Due to the fact that participants expressed themselves overtly sexist in an anonymous setting, Strait et al. (2017) assumed participants might respond less sexist in public. This in turn might support Bernotat and colleagues' (2017; 2021) assumption that social and particularly non-sexist norms could play a role HRI. The perspective that social and particularly non-sexist norms might apply to HRI is still relatively new, but it will surely gain importance, the more robots enter people's daily lives (see also Wullenkord & Eyssel, 2020a).

2 Research Aims and Overview of Experiments

Based on the literature, four main aims were followed in this PhD project.

These aims were to investigate whether

1. Attributes' and actions' gender-stereotypicality (see Bem, 1974; Berger & Krahé, 2013)
2. The speaker voice (see Van Berkum et al., 2008)
3. Participants' benevolent and hostile sexism (Glick & Fiske, 1996; German version: Eckes & Six-Materna, 1999), normative gender role orientation (Athenstaedt, 2000), motivation to control for sexist responses (Klonis et al.,2005; German version: Eyssel, 2010), and social desirability (Stöber, 2001; German version: Stöber, 1999)

would affect participants' visual attention.

4. Moreover, it was researched whether the results of Experiment 1 and Experiment 2 (see below) would be replicated when male and female robot targets were displayed (see Bernotat et al., 2017; 2021).

Taking advantage of different theoretical and methodological approaches of the fields of psycholinguistics, social psychology, and social robotics, four visual world eye tracking experiments were conducted to pursue these research aims:
In Experiment 1, gender-stereotypical <u>adverbs</u> referred to male and female <u>human</u> targets.
In Experiment 2, gender-stereotypical <u>main verbs</u> referred to male and female <u>human</u> targets.
In Experiment 3, gender-stereotypical <u>adverbs</u> referred to male and female <u>robot</u> targets.
In Experiment 4, gender-stereotypical <u>main verbs</u> referred to male and female <u>robot</u> targets.

In all experiments, the sentences were uttered by a male (vs. female) speaker voice. In addition, eye tracking data was complemented by self-report measures.

To enhance the comparability of the results, it was aimed for a high standardization across experiments. Moreover, the implementation of the visual world experiments was time-consuming and complex. The experiments had thus to be planned thoroughly, far ahead, and in conjunction to each other. Therefore, though the numeration of the experiments might suggest a sequential order, the experiments were run and analyzed partly in parallel. This might be useful to know in order to understand why ideas for future research that followed from single experiments were outlined for follow-up research in the general discussion.

3 Experiment 1 (Adverbs – Humans) – Hypotheses

The following hypotheses were derived from the literature presented to pursue the underlying research aims:

1 The adverbs' gender-stereotypicality and

2 a speaker voice whose gender matches (vs. mismatches) the adverbs' gender-stereotypicality

lead to higher log-gaze probabilities to look at a character whose gender matches (vs. mismatches) the adverbs' gender-stereotypicality[1]

Log-gaze probabilities to look at a character whose gender matches (vs. mismatches) the adverbs' gender-stereotypicality are <u>higher</u>, the higher participants' values on

3a Benevolent and hostile sexism and normative gender role orientation

and <u>lower</u>, the higher participants' values on

3b Internal and external motivation to control for sexist responses and social desirability.

[1] Log-gaze probabilities were calculated with fixations on one character relative to the other. Therefore, more looks at the character whose gender matches the adverbs' gender-stereotypicality result in fewer looks at the stereotype-inconsistent character, and vice versa (see also Arai et al., 2007).

4 Experiment 1 (Adverbs – Humans) – Method

4.1 Pretest I – Evaluation of the Attributes' Gender-Stereotypicality and Connotation

To investigate the impact of the adverbs' gender-stereotypicality on participants' anticipatory eye movements (see Hypothesis 1), Pretest I served to identify *attributes*[2] that could function as adverbs in the experimental sentences of Experiment 1 (see Section 4.6). Following Berger and Krahé (2013), attributes had to be considered gender stereotypical and as having a positive or a negative connotation in Western society in order to be used in the experimental sentences. The differentiation between positively and negatively connoted adverbs allowed to control whether the adverbs' connotation[3] affected participants' interpretation of the attributes' gender-stereotypicality. Attributes that could not be identified as stereotypically male or female were used to create the filler sentences of Experiment 1 (see Section 4.7).

4.1.1 Pretest I – Procedure

Pretest I was done in paper-pencil form. Completing the questionnaire took about 15 minutes. A set of 168 attributes in total was created. The attributes were taken either from Berger and Krahé (2013) or gained from brainstormings in seminars and lab meetings. To avoid effects of word frequency and familiarity (see e.g., Connine et al.,1990; Hyönä & Olson, 1995; Rayner & Duffy, 1986; White, 2008), attributes were selected that were known from common parlance. To keep Pretest I as short as possible for the participants, the full set of 168 attributes was split into two lists. Participants were given one of the two lists comprising half of the attributes. In a first part of the questionnaire, participants were asked to judge the attributes' gender-stereotypicality (1 = stereotypically male, 7 = stereotypically female). In a second part of the questionnaire, they evaluated the attributes' connotation (1 = positive, 7 = negative), each according to their guesses about Western Society's standards. The scale midpoint of four was considered neutral in terms of gender stereotypes and connotation, respectively (see Appendix A, p. 185 for Pretest I instructions). Although the same attributes were judged in both parts of the questionnaire, the order in which they were presented varied between the two parts. This was done to avoid sequence or memory effects. Finally, participants were asked to indicate

[2] The term *attributes* is used because adjectives and adverbs appear identical in German language. It depends on a sentence's content whether an attribute functions as an adjective or an adverb (Wöllstein-Leisten et al., 2016).

[3] The term *connotation* was considered appropriate in this context because it is defined as a word's valence and the subjective emotional arousal it may elicit. A word's connotation is culture- and language-specific (De Deyne et al.,2020; Rickheit et al.,2004).

demographics (i.e., age, gender, professional status, nationality, and native language). If needed, they received 0.5 course credits for participation.

4.1.2 Pretest I – Sample

N = 139 participants[4] (male: n = 44, female: 95; M_{age} = 27.07; SD_{age} = 10.13, age range: 19-72 years) completed the survey. Participants were recruited at Bielefeld University (students: n = 77, professionals: n = 37, undisclosed: n = 25). Most respondents were German nationals (German nationality: n = 132, other nationality: n = 6, undisclosed: n = 1). All participants indicated a good command of German language (German native speaker: n = 127, other native language: n = 11, undisclosed: n = 1).

4.1.3 Pretest I – Results

One-sample t-tests against the scale midpoint of 4 were computed to identify an attribute's gender-stereotypicality and connotation. Means statistically significant[5] below the scale midpoint indicated that the attributes were considered stereotypically male or as positively connoted in Western Society. Means statistically significant above the scale midpoint indicated that the attributes were perceived as stereotypically female or as being negatively connotated, respectively. Attributes whose means did not differ statistically significantly from the scale midpoint were identified as gender-neutral in terms of stereotypes or as neither positively nor negatively connoted, respectively[6].

In total, 62 attributes were identified as stereotypically male. 26 of them were positively connoted, 35 were negatively connoted, and one attribute was rated as neither positive nor as negative.

78 attributes were rated as stereotypically female. The majority of them – 53 attributes – were positively connoted, while 21 attributes were negatively connoted and four attributes were judged as neither positive nor as negative.

28 attributes were evaluated as gender-neutral in terms of stereotypes. 18 of them were positively connoted, 9 attributes were negatively connoted, and one attribute was judged neither as positively nor as negatively connoted. The full set of items and their ratings on gender-stereotypicality and connotation can be found in Appendix A (Table A1 – Table A9).

---

[4] In the present experiments, sample sizes were determined according to the heuristic that statistical requirements are fulfilled with a minimum of 20 participants per cell (see Field, 2013).

[5] $p < .05$. To enable a valid interpretation of p-values, effect sizes were reported complementary (see Morris & Fritz, 2013; Volker, 2006).

[6] Considering the risk of alpha inflation, in most cases, decisions about an attribute's gender-stereotypicality and connotation were based on ps < .001. In addition, decisions about an attribute's gender-stereotypicality and connotation could be confirmed by the literature (see Berger & Krahé, 2013).

4.2 Pretest II – Evaluation of Noun and Verb Phrases' Gender-Stereotypicality

To make sure that only an attribute's gender-stereotypicality affected participants' anticipatory eye movements, *nouns* and *verbs* at the beginning of the sentence had to be perceived as gender-neutral in terms of stereotypes. Pretest II thus served to identify nouns and verbs that were considered stereotypically gender-neutral in Western Society. To conduct the experimental sentences of Experiment 1, ideally not only the noun itself, but the combination of noun and verb should be considered stereotypically gender-neutral. Nouns and verbs that were not judged as neutral with regard to gender-stereotypes, were used to conduct the filler sentences of Experiment 1. Importantly, all noun-verb combinations had to be meaningful.

4.2.1 Pretest II – Material: The Noun-Verb Combinations

A priori, a set of 140 noun-verb combinations was collected. Some of the noun-verb combinations were adapted from Guerra and colleagues (2021)[7]. However, most of them were newly generated by convenience samples of university students. To avoid effects of word frequency and familiarity (e.g., Connine et al.,1990; Hyönä & Olson, 1995; Rayner & Duffy, 1986; White, 2008) nouns and verbs were selected that were frequently used in common parlance. For a better understanding of how noun-verb combinations were created, some rules of German grammar need to be explained first. In German language, the nominative is the basic form of a noun. It is either marked by a male ("der"), a female ("die"), or a neutral ("das") determiner (Wöllstein-Leisten et al., 2016). To avoid a bias due to a noun's grammatical gender, only nouns were selected that had a grammatically gender-neutral determiner, e.g., "das Paket" [the parcel]. According to linguistic standards, the combination of determiner and noun will be referred to as *noun phrase* (NP) in the following. Unlike in some languages, e.g., in French and Spanish, verbs are not gender-marked in German (Wöllstein-Leisten et al., 2016). Therefore, no grammatical form had to be considered regarding the verbs. In Pretest II, the *verb phrase* (VP) consisted of the combination of the auxiliary verb "wird" [is] (Verb 1 = V1) and a main verb (Verb 2 = V2). In German language, a combination of the auxiliary verb "wird" [is] and a main verb, results in the passive voice (Wöllstein-Leisten et al., 2016) which was intended to be used in the sentences of Experiment 1 according to Guerra, et al. (2021). To illustrate, "das Paket" [the parcel] + "wird bestellt" [is ordered] is exemplary for a noun-verb combination used in Pretest II.

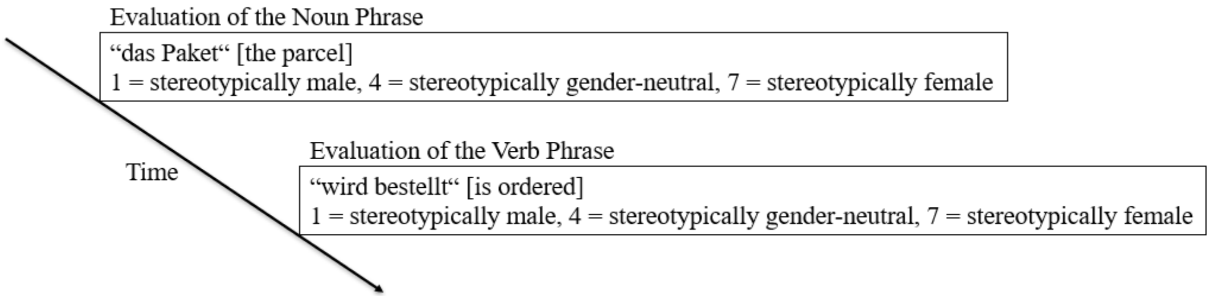
[7] The paper was based on my master thesis (Bernotat, unpublished master thesis). Thus, it was possible to create the stimuli of the present experiments according to Guerra and colleagues (2021), though their research has just recently been published.

4.2.2 Pretest II – Procedure

To keep the pretest as short as possible for the participants, the initial set of 140 noun-verb combinations (that is, a set of 140 noun phrases and 140 verb phrases) was divided into two lists of equal length. Participants were given either of the two lists which contained 70 noun and 70 verb phrases accordingly. As in Pretest I (see Section 4.1), participants were asked to estimate the noun phrases' and the verb phrases' gender-stereotypicality according to their guesses about Western Society's standards. They used bipolar 7-point Likert scales to give their responses (1 = stereotypically male, 7 = stereotypically female). The scale midpoint of four was considered gender-neutral in terms of stereotypes. To evaluate a noun phrase's perceived gender-stereotypicality independent of a verb phrase's perceived gender-stereotypicality, the questionnaire was conducted in *Unipark* (*Tivian*, former *Questback*), a tool for online-surveys. Unipark allowed to present noun and verb phrases in isolation one after another. That is, first a noun phrase was displayed on the screen whose gender-stereotypicality participants had to judge. In the following, a verb phrase was presented on the screen whose gender-stereotypicality participants rated. Then, another noun phrase was presented and so forth (see Figure 4.1). Participants were asked to judge each noun and each verb phrase independent of its antecedent (see pp. 195 for Pretest II instructions).

Figure 4.1

Schematic Depiction of the Evaluation of a Noun and a Following Verb Phrase



This procedure of evaluating noun and verb phrases independently was advantageous because it allowed both ways: If necessary, noun and verb phrases could be combined in later experimental sentences that had not been presented successively in Pretest II. Likewise, noun and verb phrases could be combined that had been presented one after the other in Pretest II (in case they had both been judged as stereotypically gender-neutral). This was the ideal case because that way, it could be considered that participants may not have been able to fully neglect an antecedent noun phrase when evaluating a verb phrase's gender-stereotypicality. Finally, participants indicated demographics (i.e., age, gender, professional status, nationality, and native language). Completing the questionnaire took about 20 minutes. If needed, participants received 0.5 course credits for taking part in Pretest II.

4.2.3 Pretest II – Sample

N = 73 participants[4] – recruited via social media platforms – took part in the questionnaire. Two participants were excluded from data analyses because they had not followed the pretest instructions. The remaining *n* = 71 participants (male: *n* = 17, female: 54; M_{age} = 29.56; SD_{age} = 13.24, age range: 18-62 years) were considered in the data analyses. The majority of them were students (students: *n* = 48, professionals: *n* = 20, undisclosed: *n* = 3) of German nationality (German nationality: *n* = 69, other nationality: *n* = 2). All participants had a good command of German language (German native speaker: *n* = 68, other native language: *n* = 3).

4.2.4 Pretest II – Results

One-sample *t*-tests against the scale midpoint of four were computed to identify a noun phrase's or a verb phrase's gender-stereotypicality. Means that did not differ statistically significantly from the scale midpoint were judged as stereotypically gender-neutral and therefore deemed suitable for the experimental sentences of Experiment 1. Noun and verb phrases whose means were statistically significant below the scale midpoint were considered stereotypically male. Noun and verb phrases whose means were statistically significant above the scale midpoint were identified as stereotypically female. Stereotypically male and female biased noun and verb phrases were used to conduct the filler sentences of Experiment 1.

In total, 50 noun phrases and 52 verb phrases were found to be stereotypically gender-neutral. 45 noun phrases and 39 verb phrases were judged as stereotypically male and 45 noun phrases and 49 verb phrases were identified as stereotypically female (see Appendix A Table A10 – Table A15 for the full set of items and the results)[8].

4.3 Pretest III – Evaluation of the Speaker Voices as Male and Female

To test the impact of the speaker voice on participants' anticipatory eye movements (see Hypothesis 2), the experimental sentences of Experiment 1 had to be recorded by a male and a female German native speaker. The major aim of Pretest III was thus to evaluate whether the speaker voices were clearly identified as male and female, respectively. Additionally, Pretest III could be regarded as a training for the speakers how to read the later experimental sentences of Experiment 1. More concretely, in case the speaker voices were deemed suitable to

[8] Considering the risk of alpha inflation, in most cases, decisions about noun and verb phrases' gender-stereotypicality were based on *p*s < .001. In addition, decisions about a noun and verb phrase's gender-stereotypicality could be confirmed by the literature (Guerra et al., 2021).

record the sentences of Experiment 1, both speakers should read the sentences in the same style. This was needed to avoid biases due to reading style and to facilitate the later editing process of the experimental sentences by keeping word region on- and offsets within-tuples and ideally also between-speakers as equal as possible (see Section 4.6.1).

4.3.1 Pretest III – Material: The Structure of the Sentences

A set of 14 sentences was created based on the results of Pretest I (Section 4.1) and Pretest II (Section 4.2). Following Guerra and colleagues (2021), German language subject-verb-object (SVO) sentences were created because this sentence structure should be maintained in Experiment 1. In line with Guerra and colleagues (2021), the sentences of Pretest III described an object[9] to be manipulated by a representative of a gender-stereotypical profession, e.g., "Das Kochfeld wird justiert von dem Mechatroniker." [The stove is adjusted by the mechatronics engineer.]. To avoid a bias due to the sentences' content, in half of the sentences, the object to be manipulated was stereotypically male, respectively female. Likewise, in half of the sentences, a stereotypically male, respectively female profession was mentioned at the end of the sentence (see Appendix A, Table A16 for the full set of sentences).

Each sentence was recorded by a male and a female speaker. To avoid a bias due to reading style and to facilitate the later editing process of the sentences (see Section 4.6.1), the speakers were trained to read the sentences clearly and with the same pronunciation, breathing times, and speech rate. More concretely, after having read the first noun phrase, e.g., "Das Kochfeld" [the stove], the speakers were told to briefly pause before reading the verb phrase, e.g., "wird justiert" [is adjusted] which was followed by a short exhale. A very short breath-in followed before reading the preposition "von" [by] that was followed by a short pause before unveiling the second noun phrase – the representative of the gender-stereotypical profession, e.g., "dem Mechatroniker" [the mechatronics engineer]. At the same time, the pauses and breaths should not be too long so that the sentences still sounded as natural as possible.

4.3.2 Pretest III – Procedure

The pretest was programmed in *Unipark* (*Tivian*, former *Questback*). Participants listened to the full set of sentences. Half of the participants first listened to each of the 14 sentences recorded by the male speaker, followed by the same set of sentences recorded by the female speaker, and vice versa. Participants were asked to carefully listen to each of the sentences. Once they had listened to a sentence, they were instructed to press a "next" button to listen to the next one. After having listened to all 14 sentences uttered by the same speaker,

[9] Grammatically, the object being manipulated was the subject of the sentence.

participants had to indicate to what extent they would agree that the speaker voice sounded interesting, female, balanced, and male (1 = not at all; 7 = very much). The items "interesting" and "balanced" served as fillers because participants were told this Pretest III was about the evaluation of speaker voices in general. The filler items were not part of further analyses. Participants were told to judge the speaker voices independent of the sentences' content. After having evaluated the set of sentences recorded by one speaker, participants were told to listen and to evaluate the same set of sentences recorded by the other speaker. Finally, they indicated demographics (i.e., gender, age, nationality, native language, and professional status). If required, they received 0.5 course credits for participation. Completing the pretest took about 10 minutes (see Appendix A, p. 209 for Pretest III instructions).

4.3.3 Pretest III – Sample

N = 78 participants[4] (male: n = 21, female: n = 56, undisclosed: n = 1; M_{age} = 24.54, SD_{age} = 8.72, age range: 17-57 years) completed Pretest III. Most of them were students (students: n = 51, employees: n = 8, undisclosed: n = 19). All participants were German nationals and had a good command of German language (German native speaker: n = 76, other native language: n = 2).

4.3.4 Pretest III – Results

Paired-sample t-tests were conducted to test whether the male voice was perceived as more male compared to the female voice and whether the female voice was perceived as more female compared to the male voice, respectively. The male voice was clearly identified as male compared to the female voice, $M_{\text{male voice}}$ = 6.58, $SD_{\text{male voice}}$ = 0.69; $M_{\text{female voice}}$ = 1.21, $SD_{\text{female voice}}$ = 0.59; $t(77)$ = -48.32, p < .001, d = 9.27. The female voice was clearly identified as female compared to the male voice, $M_{\text{female voice}}$ = 6.23, $SD_{\text{female voice}}$ = 1.16; $M_{\text{male voice}}$ = 1.13, $SD_{\text{male voice}}$ = 0.61; $t(77)$ = 32.38, p < .001, d = 5.47. Both speakers were thus suitable to record the sentences of Experiment 1.

4.4 Experiment 1 – Eligibility Requirements and Sample

Using eye tracking, comprehender characteristics such as age, gender, educational level, and linguistic and life experience have shown to impact language comprehension within only a few milliseconds (Münster & Knoeferle, 2017; 2018). To possibly avoid such biases in Experiment 1, a sample of 18- to 32-year-old German native speakers who had not been exposed to a second language before the age of six was selected. This sample was also advantageous because it has shown to integrate verbal and visual information more efficiently than other samples (see Münster & Knoeferle, 2017; 2018). To possibly control for gender effects, male and female participants were balanced across the experimental conditions.

N = 92 participants[4] recruited at Bielefeld University took part in the current Experiment 1. Six participants had to be excluded from data analyses either due to technical concerns or because it turned out they had been exposed to a second language before the age of six[10]. All of the remaining n = 86 (male: n = 39, female: 45, undisclosed: n = 2; M_{age} = 22.52, SD_{age} = 2.64, age range: 18-30 years) were German native speakers with normal or corrected-to normal vision and audition. Most of them were German nationals (German nationality: n = 83, other nationality: n = 2, undisclosed: n = 1). All participants were naïve about the purpose of Experiment 1 and ensured to have understood the experiment instructions. Accordingly, all individuals were capable of correctly answering the questions followed by the filler trials.

4.5 Experiment 1 – Design

In Experiment 1, a mixed design was realized with adverb gender-stereotypicality (male vs. female) and NP2 gender (male vs. female) as within-participants factors and with speaker voice (male vs. female) as a between-participants factor.

4.6 Experiment 1 – Verbal Stimuli: The Experimental Sentences

Following Guerra and colleagues (2021), German language subject-verb-object (SVO) sentences were used in Experiment 1. The sentences were grammatically correct whereas non-canonical in German language. Nonetheless, only SVO sentences allowed to mention an adverb followed by a representative of a gender-stereotypical profession. It was thus the only sentence structure that allowed me to investigate the impact of an adverb's gender-stereotypicality on participants' anticipatory eye movements toward a character that matched (vs. mismatched) the adverb's gender-stereotypicality in terms of actual gender and the gender-stereotypicality of the profession it represented (see Hypothesis 1). Therefore, the character that matched an adverb's gender-stereotypicality is referred to as the *gender-matching character* or *target* in the following. The character that mismatches an adverb's gender-stereotypicality is referred to as the *stereotype-inconsistent character* or *target.*

The sentences described in which manner an object was manipulated by a representative of a stereotypically male or female profession, e.g., "Das Fischfilet wird überzeugt

[10] Participants' age and native tongue were queried during the recruitment process. However, some participants denied having been exposed to a second language before the age of six and being older than 32 years when being asked during the recruitment process, but indicated a second native language or being older than 32 years when completing the questionnaire at the end of the experimental session.

gebraten von dem Mechatroniker." [The fish fillet is confidently roasted by the mechatronics engineer.].

In the experimental sentences, the NP1 (e.g., "Das Fischfilet" [The fish fillet]) was a grammatically and stereotypically gender-neutral noun in the basic form (see Pretest II, Section 4.2). To investigate the impact of the adverb's gender-stereotypicality on participants' visual attention, the adverb that was embraced by the verb phrase was identified as stereotypically male or female according to Western Society's standards. Furthermore, the adverb was judged either as positive or negative according to Western societal beliefs (see Pretest I, Section 4.1). The main verb (V2) was always a stereotypically gender-neutral verb (verbs are grammatically gender-unmarked in German, see Pretest II, Section 4.2). It was followed by the preposition "von" [by] that referred to the sentence's target (NP2). The NP2 consisted of a stereotypically male or female professional role name and its determiner used from Guerra and colleagues (2021). In this context, it needs to be explained that grammatically, the NP2 is a dative object. A male dative object in the singular form is indicated by the determiner "dem" [the]. A female dative object in the singular form is indicated by the determiner "der" [the] which is also the male determiner for nouns in the nominative, respectively, in the basic form (see Wöllstein-Leisten et al., 2016). Furthermore, in German language, female role names are commonly marked by the suffix "-in" (in some exceptions, the suffices "-mann" = male form or "-frau" = female form indicate a professional role's gender). To illustrate, "Mechaniker" refers to a male mechatronics engineer, while "Mechatroniker<u>in</u>" refers to a female mechatronics engineer. In the experimental sentences, stereotypically male professions were used in the male form. Analogously, stereotypically female professions were used in the female form.

To investigate the impact of the adverbs' gender-stereotypicality on participants' visual attention, 32 pairs of sentences, so-called *tuples*, were created. Sentences within a tuple were identical except for the adverb's gender-stereotypicality. That is, one sentence of a tuple contained a stereotypically male adverb, while the other contained a stereotypically female adverb. Adverb connotation and NP2 gender varied between tuples (see Table 4.1 for an exemplary presentation of the experimental sentences' structure; see Appendix A, Table A17 for the full set of experimental sentences of Experiment 1). Furthermore, to investigate the impact of the speaker voice on participants' visual attention (see Hypothesis 2), each sentence was recorded by a male and a female German native speaker whose voices were clearly identified as male and female, respectively (see Pretest III, Section 4.3).

Table 4.1

Schematic of the Experimental Sentences Within- and Between-Tuples

Tuple	NP1	V1	Adv.	V2	Prep.	NP2
1	Das Fischfilet	wird	**überzeugt**	gebraten	von	dem **Mechatroniker.**
	[The fish fillet	is	**confidently**	roasted	by	the **mechatronics engineer**.]
			stereotypically **male adv.**, **positive** connotation			stereotypically **male** profession
	Das Fischfilet	wird	**hingebungsvoll**	gebraten	von	dem **Mechatroniker.**
	[The fish fillet	is	**dedicatedly**	roasted	by	the **mechatronics engineer**.]
			stereotypically **female adv.**, **positive** connotation			stereotypically **male** profession
2	Das Portemonnaie	wird	**zuverlässig**	gesucht	von	der **Hausfrau**.
	[The purse	is	**reliably**	searched	by	the **housewife**.]
			stereotypically **male adv.**, **positive** connotation			stereotypically **female** profession
	Das Portemonnaie	wird	**pragmatisch**	gesucht	von	der **Hausfrau.**
	[The purse	is	**pragmatically**	searched	by	the **housewife**.]
			stereotypically **female adv.**, **positive** connotation			stereotypically **female** profession
3	Das Gift	wird	**unvorsichtig**	entsorgt	von	dem **Soldaten.**
	[The poison	is	**carelessly**	disposed	by	the **soldier**.]
			stereotypically **female adv.**, **negative** connotation			stereotypically **male** profession
	Das Gift	wird	**übervorsichtig**	entsorgt	von	dem **Soldaten**.
	[The poison	is	**overcautiously**	disposed	by	the **soldier**.]
			stereotypically **female adv.**, **negative** connotation			stereotypically **male** profession
4	Das Handy	wird	**gefühllos**	ausgeschaltet	von	der **Flugbegleiterin.**
	[The cell phone	is	**callously**	turned-off	by	the **stewardess.**]
			stereotypically **male adv.**, **negative** connotation			stereotypically **female** profession
	Das Handy	wird	**heimtückisch**	ausgeschaltet	von	der **Flugbegleiterin.**
	[The cell phone	is	**insidiously**	turned-off	by	the **stewardess.**]
			stereotypically **female adv.**, **negative** connotation			stereotypically **female** profession

Note. The original version of a set of experimental sentences and their translation to English in parentheses below. NP = Noun Phrase, V = Verb, Adv. = Adverb, Prep. = Preposition.

4.6.1 Creating and Editing the Experimental Sentences

When creating the sentences, some aspects had to be considered concerning the combination of the adverbs and the NP2 targets, the recording of the sentences, and the editing process of the later recordings. Regarding the combination of the adverbs and the NP2 targets, it was important to avoid any bias due to gender-stereotypicality or connotation. To do so, adverb gender-stereotypicality was balanced within tuples, while adverb connotation and NP2 gender were balanced between tuples (see Table 4.1; see also Appendix A, Table A17). Furthermore, because word length had shown to affect language processing (e.g., Hauk & Pulvermüller, 2004; Hyönä & Olson, 1995), the tuples were created so that the adverbs within a tuple had the same word length. Paired-*t*-tests were performed to ensure that, within-tuples, the adverbs did not differ statistically significantly in the number of *phonemes*, the smallest distinguishable units of speech used as a measure of word length (see Rickheit et al., 2004). To avoid effects of speech duration and reading style, both speakers were instructed when to take short breaks and breaths and with which speech rate and pronunciation to read the sentences similar to Pretest III (see Section 4.3). This moreover facilitated to edit the recorded sentences so that the word regions' (i.e., NP1, V1, Adv., V2, Prep., NP2) on- and offset times were as equal as possible within-tuples and ideally also between-speakers (see Appendix A, Table A17). Equal on- and offset times of word regions consequently resulted in equal duration times of word regions within-tuples which was important for later data processing and analyses. Word regions' onsets were defined as starting from voice onset timing, while word regions' offsets were counted up to voice offset timing[11]. The adverb region was extended by adding a pause of 400ms after the adverb's offset[12]. This was done to give participants enough time to process the adverb and thus to avoid interferences by other word regions as far as possible. Keeping on- and offset times within-tuples and ideally also between-speakers as equal as possible required a very precise editing and documentation process of the audio recorded sentences. If necessary, short pauses of a few milliseconds were inserted or deleted to align on- and offset times of the sentences of a tuple. The sentences were edited in *Audacity* (Version 2.1.0, *Audacity Team*). Small differences within-tuples and particularly between-speakers could not be avoided because it was crucial that the sentences still sounded as natural as possible. Therefore, time differences in the sentences within- and between-tuples were considered in later data processing and analyses (see Section 5).

4.7 Experiment 1 – Verbal Stimuli: The Filler Sentences

In addition to the 32 experimental tuples, 40 filler sentences were added to the item set. The fillers were not intended to be considered in further statistical analyses. Their main purpose was to prevent participants from guessing the hypotheses. Therefore, they had to differ from the experimental sentences in some aspects.

The filler sentences were created following Guerra and colleagues (2021). More specifically, the SVO (subject-verb-object, see Section 4.3) structure that was used in the experimental sentences was maintained. An object was described to be manipulated in a certain

[11] Voice onset timing is defined as the period between the initial burst of air and the articulation of a phoneme. Likewise, voice offset time defines the period between the end of an articulated phoneme and the burst of air that follows the articulation of this sound (see Scovel, 1998; Singh et al.,2016).

[12] Linguistic research has shown that participants need about 200ms to process verbal information (Cooper, 1974; Tanenhaus et al., 1995). Therefore, 400ms were deemed appropriate to give participants enough time to process the adverb. The pause of 400ms was exploratory pretested in the lab. None of the lab members and at a later point, none of the participants indicated to have noticed the pause.

manner by a representative of a gender-stereotypical profession. Unlike the experimental sentences, the filler sentences were not arranged in tuples because the adverbs' gender-stereotypicality had not to be manipulated in the fillers.

Accordingly, the gender-stereotypicality of the fillers' word regions differed from those of the experimental sentences: In the filler sentences, the object to be manipulated (NP1) was represented by a grammatically gender-neutral noun that was judged as stereotypically male or female. It was followed by a stereotypically male or female main verb (see Section 4.2). The adverb was neutral in terms of gender stereotypes. I would have selected adverbs that were also judged neither as positively nor as negatively connoted. This however was not possible because, according to Pretest I (Section 4.1), only one adverb was judged as neutral in terms of both, gender-stereotypicality and connotation (see Appendix A, Table A3). Therefore, stereotypical gender-neutral adverbs were used with either a positive or a negative connotation. The NP2 target was a representative of a gender-stereotypical profession. Unlike the experimental sentences, the professional roles were used in the male or in the female form. That way, a NP2 target's grammatical gender and its profession's gender-stereotypicality could mismatch (e.g., "Mechatronikerin" [mechatronics engineer = stereotypically male profession used in the female form], "Hausmann" [househusband = stereotypically female profession used in the male form]).

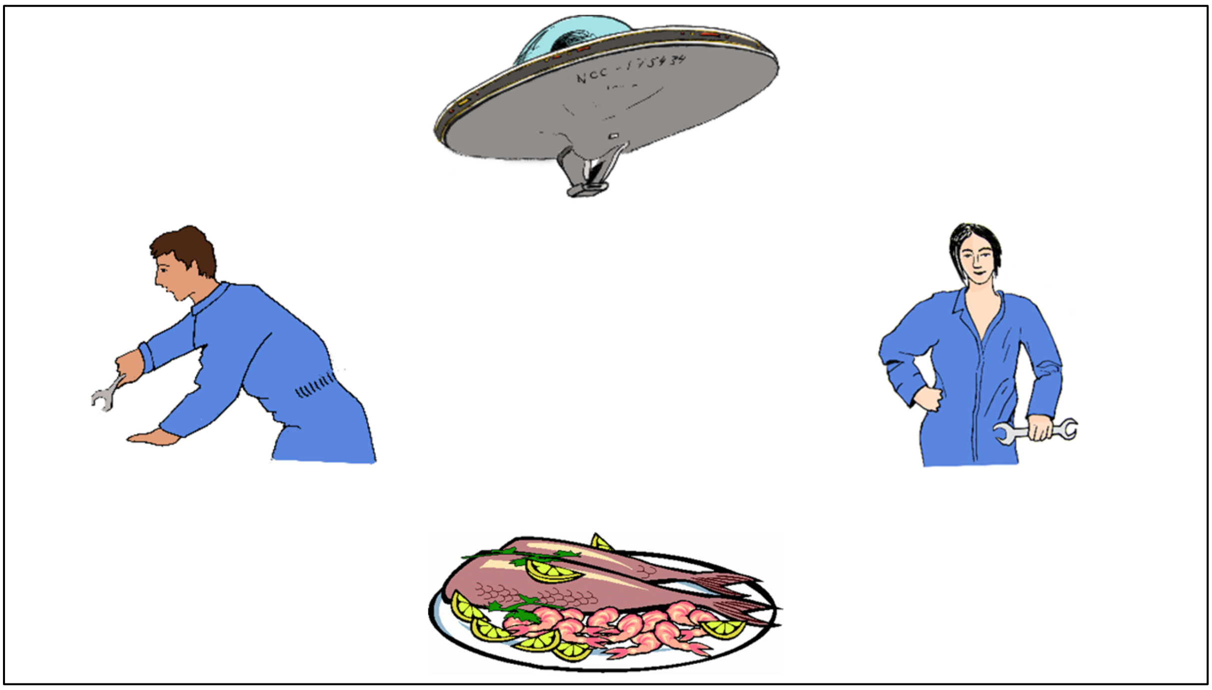
Although the fillers were not needed for hypotheses testing, they had to be created carefully in order not to bias participants in any regard. Therefore, the nouns' and verbs' gender-stereotypicality, the adverbs' connotation, the NP2 target's gender-stereotypicality, and the NP2 target's actual gender were as balanced as possible. Because the fillers were not included in later statistical analyses, on- and offsets and durations of single word regions did not have to be edited and documented. Nonetheless, to avoid any bias due to reading style, the speakers read the fillers in the same manner as the experimental sentences (see Section 4.6.1).

4.8 Experiment 1 – Visual Stimuli

According to the visual world paradigm (Allopenna et al., 1998; Cooper, 1974; Magnuson et al., 1999; Tanenhaus et al., 1995), the sentences were accompanied by a visual scene that represented the respective verbal input. Following Guerra and colleagues (2021), the visual scenes contained colored drawings of the NP1 object, a male and a female character that represented the mentioned gender-stereotypical profession, and an unrelated distractor object. The visual scenes were always arranged in the same manner, no matter whether they related to the experimental sentences or to the filler sentences. Concerning the experimental sentences, the same visual scene referred to both sentences of a tuple (see Figure 4.2 for an exemplary visual scene).

Figure 4.2

The Visual Scene Referring to Tuple 1 in Table 4.1 (see also Appendix A, Table A17)



Note. The NP1 object (the fish fillet) is depicted amongst the NP2 characters – a male and a female mechatronics engineer – and an unrelated distractor object.

Creating the visual scenes, it was important that characteristics of single stimuli did not distract participants' visual attention from the sentences' content.

Therefore, all stimuli were colored drawings that matched in style and size (see Guerra et al., 2021). Such a match was important to prevent *pop out effects*[13] of single stimuli (see e.g., Connor et al.,2004; Treisman & Gormican, 1988; see Wolfe & Horowitz, 2004 for attributes that guide visual attention). Keeping objects and characters constant in size within and across visual scenes furthermore enabled to define areas of interest (AIs) of the same size. AIs were needed to capture participants' fixations on a visual stimulus over time (see Section 5.1). Because participants' fixations on the characters were in focus of investigation, it was especially important that the AIs that framed the male and the female character had exactly the same size for valid data analyses (Holmquist et al.,2011).

Regarding the stimuli's array, the characters within a visual scene were always displayed horizontally, while the NP1 and the distractor objects were always depicted vertically (see Figure 4.2). To avoid fixation preferences due to the stimuli's position (see e.g., Campana & Casco, 2009; Geyer et al.,2007; Maljkovic & Nakayama, 1996), for half of the scenes, the

[13] A pop-out effect occurs when a stimulus is different from its surrounding and thus attracts participants' visual attention (see e.g., Connor et al.,2004; Treisman & Gormican, 1988).

NP1 object was depicted above the distractor object, and vice versa. Likewise, for half of the scenes, the male character was left to the female character, and vice versa (see also Guerra et al., 2021). To furthermore avoid *effects of cueing*, characters that were depicted as performing an action were directed toward the outer edge of the screen. More precisely, if a character that performed an action would have been directed to the middle of the screen, it might have 'pointed' toward the other character. This might have affected participants' visual attention (see e.g., Frischen et al.,2007; Posner, 1980; see Appendix A, pp. 223 for the visual scenes).

4.9 Experiment 1 – Questionnaire

A paper-pencil questionnaire was used to assess participants' sexist attitudes, normative gender role orientation, motivation to control for sexist responses, social desirability, demographics (i.e., age, gender, nationality, native language), and their guesses about the purpose of the experiment. Participants expressed their agreement to the presented statements using 7-point Likert scales with high values representing high agreement with the respective statement. Items were recoded if necessary. Internal consistencies (Cronbach's α; Cronbach, 1951) for the respective constructs were calculated for each experimental condition – that is, per male and female speaker voice. Internal consistencies were moderate to high (see Nunnally, 1978[14]) In the following, mean scores of each construct were computed. High values indicate high endorsement of the respective construct (see Appendix B, Table B1 for internal consistencies and mean scores per speaker voice condition). Instead of assessing each construct successively, the items of the constructs were presented in a randomized order. This order was the same for all participants. An overview of the measured constructs is given in the following. The full set of items including their original wording in German and their translation to English is listed in Appendix A (Table A19 – Table A22).

4.9.1 Ambivalent Sexism Inventory (ASI; Glick & Fiske, 1996, German Version: Eckes & Six-Materna, 1999)

The instrument contained a total set of 22 items with eleven items assessing participants' endorsement of benevolent sexism (e.g., "No matter how accomplished he is, a man is not truly complete as a person unless the love of a woman.") and eleven items to measure participants' endorsement of hostile sexism (e.g., "Women are too easily offended.").

[14] Nunnally (1978) postulated an alpha of 0.7 to be acceptable. However, Schmitt (1996) and Cortina (1993) noted to use such cut-offs with caution because the interpretation of alpha depends on the number of items within a scale.

4.9.2 Normative Gender Role Orientation (NGRO; Athenstaedt, 2000)

This instrument comprised 29 items to measure whether participants conform to rather egalitarian or traditional gender roles (e.g., “Women are less interested in politics than men.”). This instrument was conducted by an Austrian author. The items’ wording was thus used as in Guerra and colleagues (2021) who had adapted some items to local participants’ parlance.

4.9.3 Social Desirability Scale (SDS-17; Stöber, 2001; German Version: Stöber, 1999)

17 items assessed participants’ tendency to respond socially desirable (e.g., “During arguments I always stay objective and matter-of-fact.”).

4.9.4 Motivation to Control for Sexist Responses (MCSR; Klonis et al.,2005; German Version: Eyssel, 2010)

A set of 20 items were used to assess participants’ internal and external motivation to respond in a non-sexist manner. Thus, ten items measured participants’ internal motivation to control for sexist responses (e.g., “I attempt to act in nonsexist ways toward women because it is personally important to me.”). Likewise, ten items tapped their external motivation to control for sexist responses (e.g., “Because of today’s PC (politically correct) standards I try to appear nonsexist toward women.”).

4.10 Experiment 1 – Procedure

Participants were recruited on a university campus. They were told that we[39] would conduct an eye tracking experiment on language processing which would be completed by a short questionnaire. Participation requirements were outlined if people signaled interest to take part in the experiment (see Section 4.4). After being welcomed to the lab, participants were introduced to the procedure of the experiment and to the eye tracking method. Furthermore, they were informed in oral and written form about data security and their right to end the experimental session any time without giving reasons. Then, they provided informed consent[15]. As part of a set of similar eye tracking experiments, Experiment 1 was approved by University of Bielefeld’s Ethics Committee (Approval No. 2016 – 042).

After having agreed to have their data recorded, participants were seated approximately 70cm from a color monitor (22-inch color monitor, 1680x1050px). Their eye movements were recorded by an *EyeLink 1000* (*SR Research*) desktop head-stabilized tracker with a sampling rate of 1000Hz. The experiment was programed and eye movements recorded using *Experiment Builder* (*EyeLink* software, *SR Research*). Eye tracking data were processed with

[15] Written consent forms and information sheets were in line with the guidelines for data security.

Data Viewer (*EyeLink* software, *SR Research*). Viewing was binocular, however only the right eye was tracked[16]. Each session began with a 9-point calibration procedure to ensure that participants could be calibrated with a tracking accuracy below 0.50°[17]. Then, participants read the experiment instructions (Guerra et al., 2021, adapted) which were shown on the screen (Times New Roman, black font, 12pt., see Appendix A, pp. 242 for experiment instructions). Participants' task was to pay attention to the sentences and to the related visual scenes. The sentences were presented using two loud speakers in normal conversation volume (~ 65db), while the visual scenes were simultaneously displayed on the screen. To ensure that participants would pay close attention to the sentences and to the visual scenes, they were informed that some trials were followed by questions concerning the sentences or the visual scenes.

After participants had affirmed that they had understood the task, a recalibration was done before the first trial was initiated. Each session consisted of 72 trials in total. 32 of the trials were experimental trials and 40 of the trials were fillers that were intended to prevent participants from guessing the hypotheses.

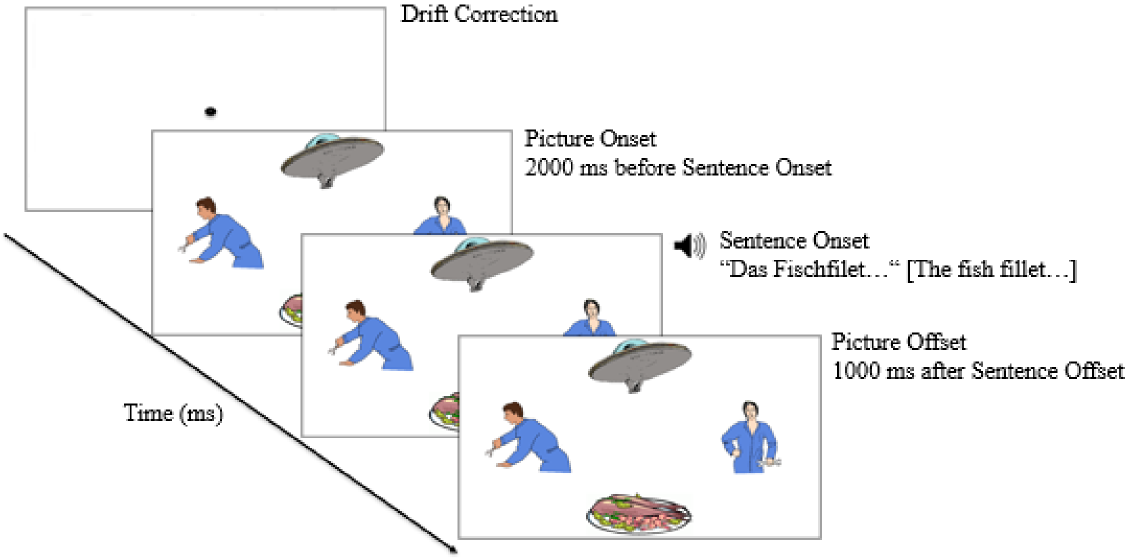
Basically, each trial proceeded as follows (see also Figure 4.3): A drift correction was initiated to validate the calibration (recalibrations were done whenever tracking accuracy was above 0.50°). Drift corrections were marked by a black fixation dot in the middle of the screen. Once participants fixated the dot and tracking accuracy was below 0.50°, the experimenter initiated the trial. Then, the visual scene was displayed for 2000ms before sentence onset. This gave participants time to explore the screen before listening to the sentence (see Huettig & McQueen, 2007; see Huettig et al.,2011b for a review). The visual scene remained until 1000ms after the sentence had been presented (see Appendix A, Table A17 for the exact duration times of each experimental sentence which is marked by its NP2 offset). To make sure that participants indeed paid attention to the sentences and to the visual scenes, the filler trials were followed by a question that either related to the content of the sentences or to the pictures. The question was displayed until participants had pressed a "yes-" or "no"-button (*Cedrus* Response Pad, 8-buttons, large). The "yes" button was on the right hand, the "no"-button was on the left hand. To avoid any bias due to handedness (see Shen & Franz, 2005), for half of the questions the correct answer was "yes", respectively "no". Half of the questions concerned the content of the sentences and the visual scenes, respectively (see Appendix A, Table A18).

[16] For most people, the right eye is the dominant eye (Bourassa et al., 1996; Erdogan et al.,2002). To standardize the tracking method, only the right eye was tracked because it was likely to yield most valid data.

[17] As in the current set of experiments, the stimuli in visual world experiments are commonly quite large and areas of interest are larger than the stimuli (see Section 5.1). I thus followed common recommendations for visual world experiments to consider a tracking accuracy below 0.5 as fairly good (see also Kröger et al., 2018).

Figure 4.3

Schematic of an Experimental Trial



After half of the trials, participants were offered to take a short break of no longer than a minute. The eye tracking procedure took about 20 minutes and was followed by a questionnaire which took about 15 minutes (see Section 4.9). For reasons of discretion, participants completed the questionnaire in a side room while the experimenter stayed in the eye tracking room. The experimenter was either a male or a female student assistant, but none of the speakers who had recorded the sentences because this may have biased participants' visual attention as well as their questionnaire responses. After having filled in the questionnaire, participants were informed about the purpose of the experiment and were thanked with chocolate and 1.0 credit if needed. In addition, they were offered to participate in a draw of three *Best-Choice* vouchers á 10€.

4.10.1 Sequence of Experimental Trials and Filler Trials

To fully explain the procedure of Experiment 1, two details need to be pointed out: First, the experimental sentences of a tuple were assigned to one of two lists (see also Arai et al., 2007; Pollatsek & Well, 1995). Participants were confronted with the experimental sentences of either the two lists and with the full set of filler sentences. This resulted in the 32 experimental trials and the 40 filler trials per session mentioned in Section 4.10. Second, experimental trials and filler trials were presented in a so-called *pseudo-randomized* order. More precisely, to possibly avoid order effects, the experimental sentences and the fillers within each list were randomized so that each participant saw the items in a different order. However, this randomization was done with two constraints: First, the first trial of an experimental session was always a filler trial. This way, participants could familiarize with the eye tracking method and make sure that they had understood the task. Second, to alternate experimental and filler trials, no more than two experimental trials were presented in a row. Due to these two constraints in the randomization process, a *pseudo-randomization* was realized.

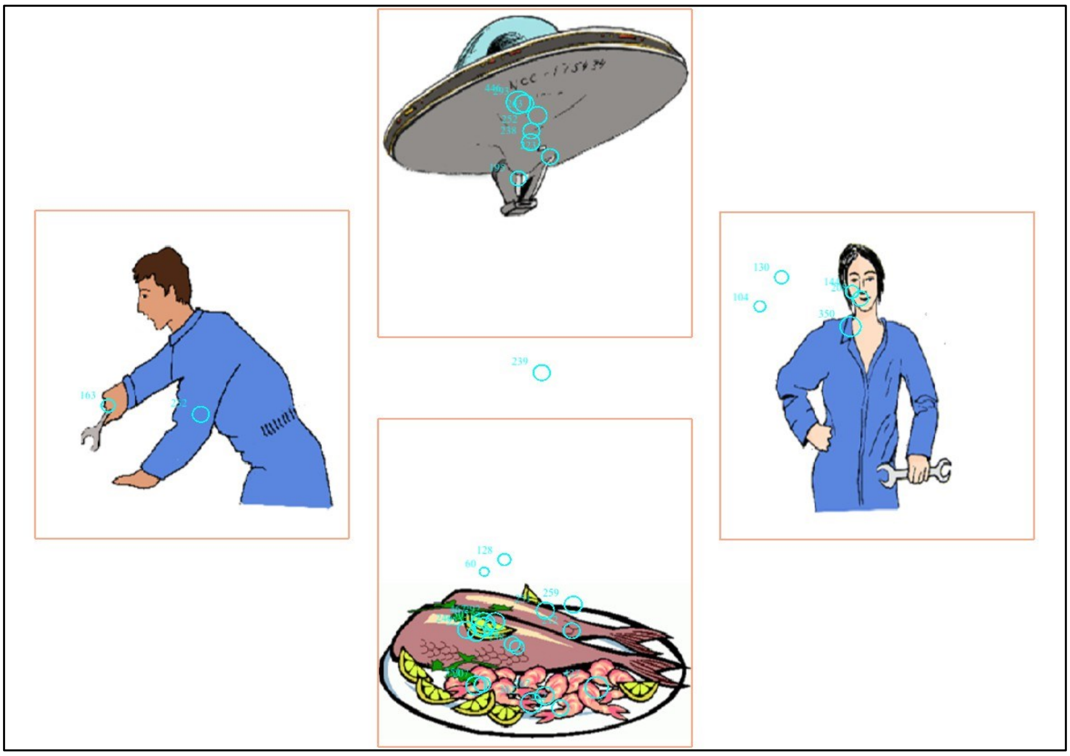
5 Experiment 1 (Adverbs – Humans) – Results

5.1 Experiment 1 – The Steps of Data Processing

Some preparations were needed before data could be analyzed. To capture all fixations[18] on the male and the female character as well as on the NP1 object and the distractor object, four areas of interest (AIs) had to be predefined (see Figure 5.1). Because the definition of AIs was required for data collection, it marked strictly speaking the *first step of data processing*. AIs were defined using *Data Viewer* (*EyeLink* software, *SR Research*). All AIs were rectangular shaped. This allowed to keep them equal in size (338px x 303px) which in turn enhanced the generalizability and validity of data analyses (see Holmquist et al.,2011).

Figure 5.1

The Set of AIs That Framed the Characters and the Objects (Rectangular Frames)[19]



[18] Fixations occur between eye movements, so-called *saccades*. During fixations, the eyes remain relatively still for about 200ms – 300ms (Rayner, 1998), except for some micro saccades and tremors (see Holmquist et al.,2011).

[19] This picture was taken from *Data Viewer* to illustrate the AIs. The turquoise circles and the numbers above indicate a participant's fixations and their durations (in ms) during the trial in which this picture was displayed. The AI frames, the turquoise circles, and the numbers were not displayed during the experiment.

Fixations during the filler trials as well as fixations on the NP1 and the distractor object only served to test whether participants had paid attention to the entire sentences and to all entities depicted. They were not relevant to test the hypotheses and were thus not part of further data processing and analyses. All fixations on the AIs that framed the male and the female character were processed as described in the following.

Log-transformed fixation proportions (also called *log-ratios*) were the basis to investigate participants' eye movements toward the male and the female character. They were computed in a *second step of data processing* (see Section 5.1.1). To do so, fixations captured by the AIs that framed the characters were used to calculate fixation probabilities on the male character relative to fixation probabilities on the female character over time. The resulting fixation proportions were then log-transformed (see Arai et al.,2007). That way, log-ratios were computed per adverb gender-stereotypicality, per speaker voice, per NP2 gender, (and per adverb connotation for exploratory analyses).

Using log-ratios per male and per female adverb gender-stereotypicality and NP2 gender, time course graphs were created in a *third step of data processing* (see Section 5.1.2). Time course graphs allowed for a rough descriptive inspection of participants' fixation patterns over the course of the entire sentence.

To prepare data for inferential statistical analyses and thus, being able to investigate the effect of adverb gender-stereotypicality, speaker voice, and NP2 gender more accurately, a *fourth step of data processing* was needed: The adverb and the NP2 region as the two word regions of interest had to be predefined exactly. Inferential analyses of the adverb region were needed to test the hypotheses (see Section 5.2). Inferential analyses of the NP2 region served as a manipulation check to investigate whether participants fixated the NP2 target after it had been mentioned, independent of the speaker voice which would indicate that participants could link the sentences' content to the displayed characters.

To finally do inferential analyses per word region, a *fifth* and last *step of data processing* was crucial: The sum of fixations each within the predefined adverb region and within the predefined NP2 region had to be aggregated, restructured, and log-transformed by experimental condition once by participants (F_1) and once by items (F_2)[20] (see Arai et al., 2007; Pollatsek & Well, 1995; Section 5.1.4). Analyzing data by participants and by items, each experimental factor (adverb gender-stereotypicality, speaker voice, NP2 gender) was

[20] In this context, each single sentence of a tuple was considered an item.

once treated as a within- and once as a between-participants/items factor. This way, an imbalance due to the fact that in mixed designs the within-factor naturally gains more data points and is thus more precisely measured than the between-factor could be compensated.

5.1.1 Experiment 1 – Step 2 of Data Processing: Calculating Log-Transformed Fixation Proportions

To analyze the effects of the experimental factors on participants' visual attention toward the male and the female character over time, log-ratios were calculated per adverb gender-stereotypicality, speaker voice, NP2 gender, (and adverb connotation for exploratory analyses).

For a better understanding, the formula below illustrates how log-ratios were calculated, using stereotypically male adverbs as an example:

$$\text{Log-Ratios for Male Adverbs} = \frac{\text{Ln}(\text{Sum of Fixations on the Male Character} + 0.5)}{\text{Ln}(\text{Sum of Fixations on the Female Character} + 0.5)}\,.$$

Log-ratios per experimental condition could be explored over the course of the entire sentence or solely within a specific word region. The principle of calculating log-ratios was always the same: First, all fixations[21] on the male and on the female character were each captured in time-bins of 20ms[22] per participant and per item. To capture fixations over time, a new variable, named *code*, was created. Scrutinizing each fixation across items and participants per time-bin, the code variable captured whether an adverb was stereotypically male or female, whether the male or the female character was depicted left or right, and whether the left or the right AI was fixated. That way, the code variable documented whether the male or the female character was fixated within a specific time-bin. Then, fixations on the male and the female character were summed each over items and over participants per experimental condition (e.g., per adverb gender-stereotypicality, see formula above). The sum of fixations on the male character was then divided by the sum of fixations on the female character per

[21] All fixations on the characters were considered in the data analyses to validly reflect participants' gaze patterns. Some authors recommend to exclude fixations below 80ms and above 1200ms from data analyses because these authors do not consider these fixations informative (e.g., Sturt et al.,2010; Rayner, 1998). However, I decided to consider all fixations because short fixations of less than 100ms duration likely occur in a pre-attentive stage of object localization (Velichkovsky et al.,2000; Velichkovsky et al.,2003). Fixations above 1000ms of duration were considered informative because they might reflect participants' difficulty to link an attribute to a gender-stereotypical character due to prevailing gender stereotypes (fixations above 1000ms rarely occurred in the data).

[22] The smaller time-bins, the more detailed data analyses are possible.

time-bin of 20ms. Since ratios of zero are undefined as are logs of zero (Arai et al., 2007, Field, 2013), a constant of 0.5 was added to each fixation sum. Finally, the ratio between the sum of fixations on the male character + 0.5 relative to the sum of fixations on the female character + 0.5 per time-bin was then log-transformed. Positive log-ratios thus indicate that the male character was fixated more frequently within a specific time-bin than the female character. Reversely, negative log-ratios indicate that the female character was favored over the male character within a specific time-bin. Log-ratios of zero imply that both characters were fixated equally likely.

Compared to standard probabilities, log-ratios were advantageous in two ways: First, fixation proportions on the male and the female character are not linearly independent of one another. That is, more fixations on the male character imply less fixations on the female character at the same time, and vice versa. Second, unlike standard probabilities that range between 0 and 1, log-ratios can take indefinite positive and indefinite negative values (Arai et al., 2007). They are therefore more likely to yield normally distributed data and to meet important requirements for parametric testing (e.g., to correct for unequal variances and a lack of linearity, Arai et al., 2007; Tabachick & Fidell, 2007).

5.1.2 Experiment 1 – Step 3 of Data Processing: Creating Time Course Graphs

Using log-ratios for stereotypically male and female adverbs that refer to either stereotypically male or female NP2 targets, time course graphs were created to descriptively explore time-lined fixation patterns on the male and the female character from sentence onset to its offset (see Figure 5.2 and Figure 5.3). To roughly assign gaze patterns to word regions, mean on- and offset times of the adverb and the NP2 region were inserted into the graphs. Mean adverb and NP2 on- and offsets were calculated using all fixations over participants and items (see Table 5.1; see also Appendix A, Table A17 for on- and offsets). Because on- and offset times of the adverb and the NP2 region differed between items, the time course graphs reach until 8640ms to include the latest NP2 offset at 8637ms and thus to display all fixations over the course of the entire sentence.

Table 5.1

Earliest, Latest, and Mean On- and Offset Times (ms) of the Adverb and the NP2 Region

Word Region	Earliest	Latest	*M*	*SD*
Adverb Onset	1913	2647	2219.46	171.88
Adverb Offset	2492	4098	3169.32	309.55
NP2 Onset	4009	6202	5163.49	399.67
NP2 Offset	5036	8637	6376.58	628.21

Note. N = 28823 (sum of all fixations over participants and items).

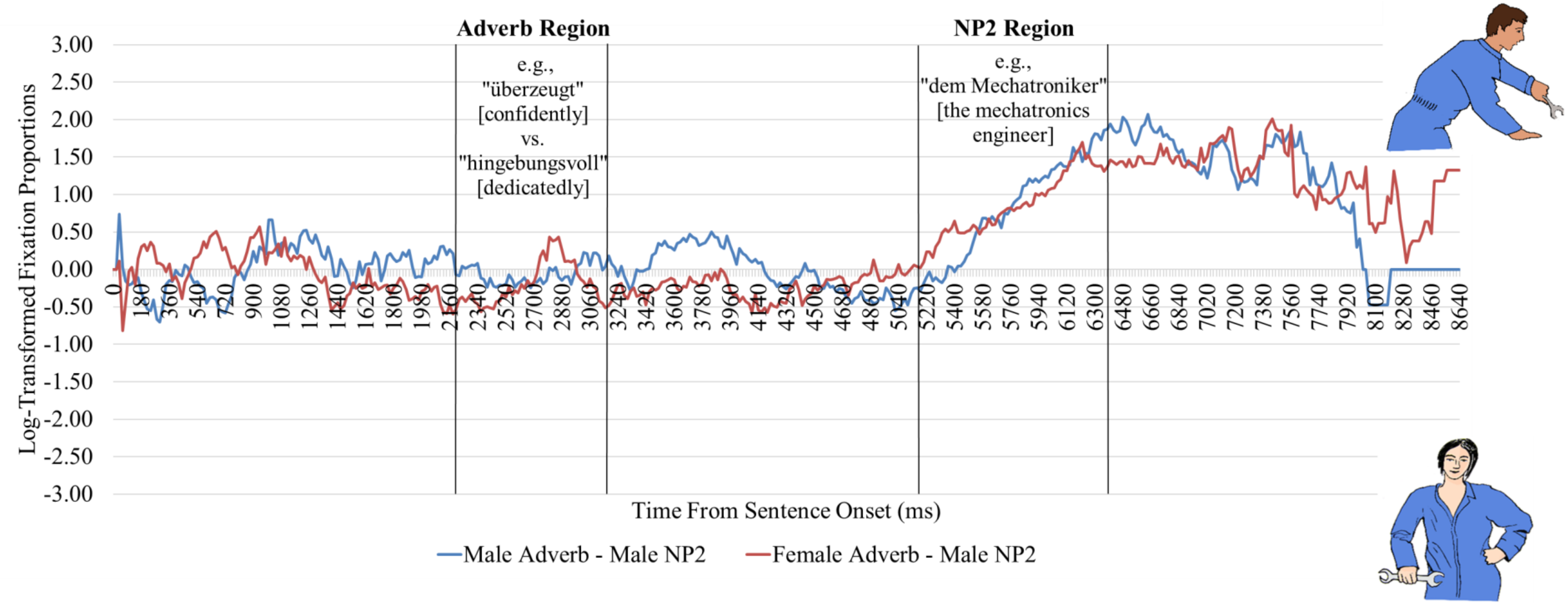
In time course graphs that visualize log-ratios for sentences that ended on a stereotypically male NP2 target (Figure 5.2), there was seemingly a slight tendency to look at the character whose gender matched the adverbs' gender-stereotypicality from about 600ms before adverb mean onset up to about 300ms after. Exploring the NP2 region, log-ratios were positive from mean NP2 onset on with values up to about 2.00 which indicates that the NP2 target was looked at when it was explicitly named at the end of the sentence.

In time course graphs visualizing log-ratios for sentences ending on a stereotypically female NP2 target (see Figure 5.3), there was also a tendency to inspect the gender-matching character about 600ms before mean adverb onset up to about 300ms after. This tendency was more pronounced as for sentences that ended on a male NP2 target. Regarding the NP2 region, the NP2 target was looked at even about 150ms before mean NP2 onset. Log-ratios within the NP2 reached values up to about -2.50. Compared to values of about 2.00 within the NP2 region when the NP2 target was male, there might be a tendency to more likely look at the NP2 target when its gender was female (vs. male).

To sum up, time course graphs for sentences ending on a male and a female NP2 target suggest that there was only a small effect of adverb gender-stereotypicality, if at all. Naming of the NP2 target seemed to have attracted participants' visual attention regardless of whether its gender was male or female. Nevertheless, inferential analyses of the adverb and the NP2 region are needed to confirm descriptive time course graphs.

Figure 5.2

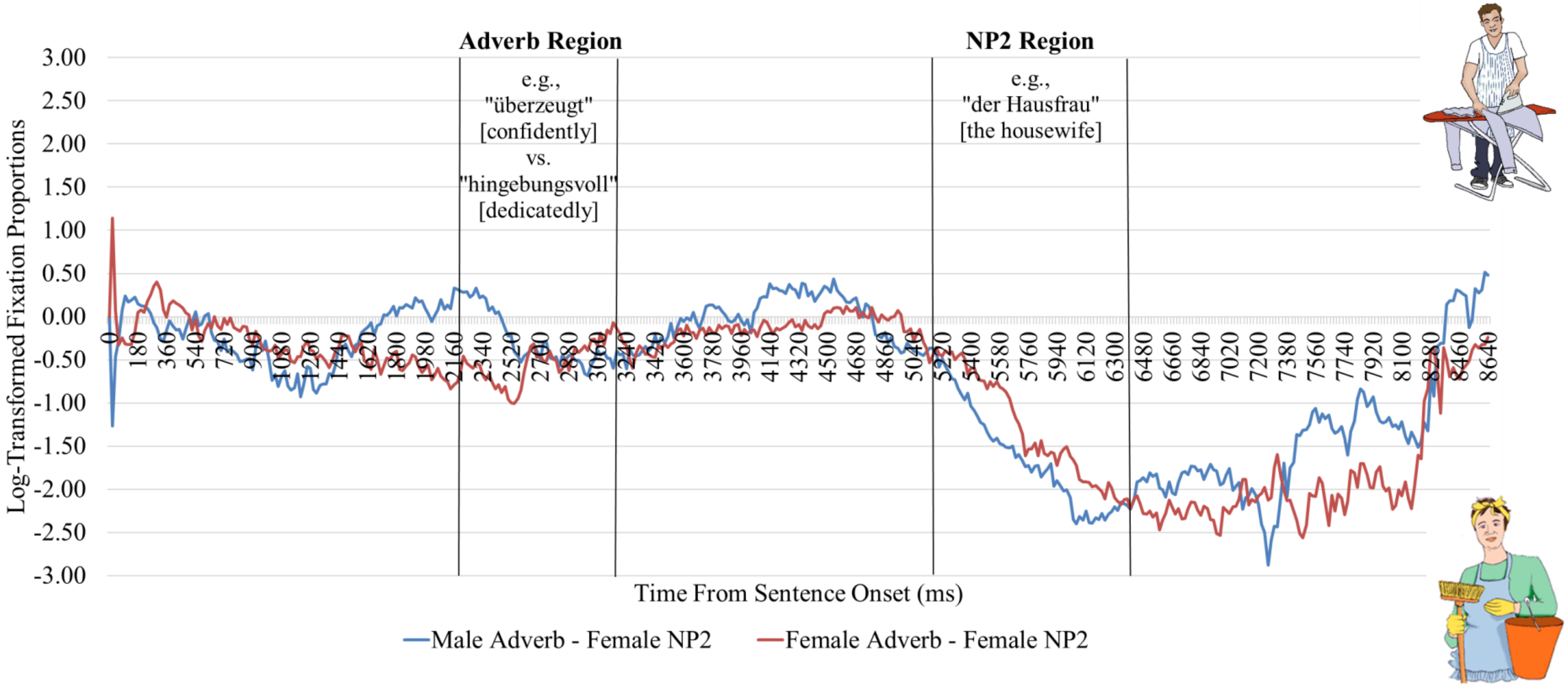
Time Course Graph for Sentences With a Stereotypically Male vs. Female Adverb Referring to a Stereotypically Male NP2 Target



Note. Time course graph depicting changes in log-transformed fixation proportions to the male (positive values) and to the female character (negative values) when encountering sentences with a stereotypically male (blue line) vs. female (red line) adverb referring to a stereotypically male NP2 target. Mean on- and offset times of the adverb region (1. and 2. vertical line) and mean on- and offset times of the NP2 region (3. and 4. vertical line) are indicated (see Table 5.1).

Figure 5.3

Time Course Graph for Sentences With a Stereotypically Male vs. Female Adverb Referring to a Stereotypically Female NP2 Target



Note. Time course graph depicting changes in log-transformed fixation proportions to the male (positive values) and to the female character (negative values) when encountering sentences with a stereotypically male (blue line) vs. female (red line) adverb referring to a stereotypically female NP2 target. Mean on- and offset times of the adverb region (1. and 2. vertical line) and mean on- and offset times of the NP2 region (3. and 4. vertical line) are indicated (see Table 5.1).

5.1.3 Experiment 1 – Step 4 of Data Processing: Defining Word Regions

Word regions were defined from the onset of the respective word to the onset of the following word (see also Arai et al., 2007; Knoeferle et al., 2005; Kröger et al.,2018). Therefore, the adverb region ranged from an item's adverb onset up to the item's main verb onset. The NP2 region was defined from NP2 onset up to NP2 offset (because this marked the sentence offset). To capture all fixations that occurred within the predefined adverb region and within the predefined NP2 region, adverb durations and NP2 durations had to be calculated for each single item.

To calculate an item's adverb duration, the onset of the item's adverb was subtracted from the item's main verb onset. The first sentence of tuple 1 recorded by a male speaker ("Das Xylophon wird sachlich gespielt von dem Bauarbeiter." [The xylophone is objectively played by the construction worker.], see Appendix A, Table A17) serves to illustrate the calculation of the item's adverb duration: For this item, the onset of the adverb "sachlich" [objectively] was at 2156ms. The onset of the main verb "gespielt" [played] was at 3221ms. Subtracting the adverb onset of 2156ms from the main verb onset of 3221ms, the adverb duration of this specific item was 1065ms. Adverb durations of all other items were calculated in the same manner. Durations of the NP2 region were computed analogously: An item's NP2 onset was subtracted from the item's NP2 offset. This was done for all items. Then, the longest adverb and NP2 region were identified across items. They marked the total length of the respective word region. That way, the duration of the adverb duration was 2130ms; the NP2 duration was 2461ms. For further data processing, fixations on the male and the female character that occurred within the respective word region were summed per item and per participant in time-bins of 20ms. To illustrate, within the adverb region, fixations were counted that occurred between an individual item's adverb onset and 2130ms (rounded to 2140ms because time-bins of 20ms were used) after. Analogously, within the NP2 region, fixations were summed that occurred between an individual item's NP2 onset and 2461ms (rounded to 2480ms) after.

5.1.4 Experiment 1 – Step 5 of Data Processing: Aggregating, Restructuring, and Log-Transforming Data by Participants and by Items

To finally prepare data for inferential statistical analyses, the sum of fixations on the male and the female character across participants and items within the predefined adverb and NP2 region needed to be aggregated and restructured by participants and by items (see Section 5.1). To better understand the aggregation and restructuring processes, one needs to bear in mind that adverb gender-stereotypicality, adverb connotation, and NP2 gender varied within-participants, but between-items. Speaker voice varied between-participants, but within-items (see Section 4.5). Therefore, data processing by participants and by items resulted in slightly different outcome variables, although aggregation and restructuring of the data were similar:

To investigate the effects of adverb gender-stereotypicality and speaker voice on participants' eye movements, the sum of fixations within the predefined adverb region was aggregated by the code variable (the variable that defined whether the male or the female character was fixated, see Section 5.1.1), adverb gender-stereotypicality, and speaker voice. This was done by participants and by items. Then, data was restructured. The restructuring process resulted in four variables each by participants and by items. Each of them counted the sum of fixations within the adverb region per within-participants/within-items factor and per character fixated. To illustrate, by participants, one variable counted the sum of fixations on the male character, the other variable counted the sum of fixations on the female character for stereotypically male adverbs. Likewise, one variable counted the sum of fixations on the male character, the other variable counted the sum of fixations on the female character for stereotypically female adverbs. Then, log-gaze probabilities for the male character were calculated relative to log-gaze probabilities for the female character per within-participants/within-items factor (see formula, Section 5.1.1). That way, by participants, one variable encompassed log-ratios for stereotypically male adverbs, the other documented log-ratios for stereotypically female adverbs per participant. By items, one variable captured log-ratios for the male speaker voice, while the other captured log-ratios for the female speaker voice per item. Data processing of the sum of fixations within the NP2 region was done analogously, but with NP2 gender as aggregation and outcome variable.

To explore whether adverb connotation had affected participants' fixation patterns, data processing was similar to that of the calculation of log-ratios for stereotypically male and stereotypically female adverbs by participants. The only difference was that the sum of fixations within the adverb region was additionally aggregated and restructured by adverb connotation. Finally, log-ratios were calculated separately for: 1. stereotypically male positively connoted adverbs, 2. stereotypically male negatively connoted adverbs, 3. stereotypically female positively connoted adverbs, and 4. stereotypically female negatively connoted adverbs. Because adverb connotation was a between-items factor, no aggregation by adverb connotation was needed by items.

Each aggregation and restructuring process of the data resulted in a separate data set. Therefore, to include all log-ratios within the adverb and the NP2 region in one statistical model, all data sets that were restructured by participants were merged. Likewise, all data sets that were processed by items were merged, too. Data processed by participants were furthermore merged with participants' questionnaire data. This was crucial to investigate the effects of participants' endorsement of benevolent and hostile sexism, normative gender role orientation, motivation to control for sexist responses, and social desirability on their fixation patterns (see Section 5.2)[23].

5.2 Experiment 1 – Testing of the Experimental Hypotheses

The adverbs' gender-stereotypicality was hypothesized to lead to fixations on the character whose gender matched (vs. mismatched) the adverbs' gender-stereotypicality (Hypothesis 1). The gender-matching (vs. mismatching) character was hypothesized to be more likely fixated when the adverbs were uttered by a speaker whose gender matched (vs. mismatched) their gender-stereotypicality (Hypothesis 2). Furthermore, participants were expected to more likely fixate the gender-matching character, the higher their values on benevolent and hostile sexism and normative gender-role orientation (Hypothesis 3a). They were expected to less likely fixate the gender-matching character which would imply more looks at the stereotype-inconsistent character, the higher their endorsement of internal and external motivation to control for sexist responses and social desirability (Hypothesis 3b).

To test Hypothesis 1 and Hypothesis 2, repeated measures multivariate analyses of variance (MANOVAs) were performed by participants (F_1) and by items (F_2). Log-ratios during the adverb region were investigated as a function of adverb gender-stereotypicality and

[23] To avoid multiple data sets and analyses by participants and by items, some authors vote for the use of mixed-effects models (see Baayen et al.,2008; Bates et al.,2018). These however bear the risk of uninterpretable models due to overparameterization (Bates et al., 2018).

speaker voice[24]. To test Hypothesis 3a and Hypothesis 3b, a multivariate analysis of covariance (MANCOVA) was performed by participants. Log-ratios within the adverb region were investigated as a function of adverb-gender-stereotypicality and speaker voice, while the effects of participants' endorsement of benevolent and hostile sexism, normative gender role orientation, internal and external motivation to control for sexist responses, and social desirability were considered as covariates[25].

For a better understanding of the potential role of the covariates, Pearson correlations between the experimental factors and the covariates that turned out to statistically significantly impact participants' eye gazes were performed (see Appendix B, Table B1 for mean values on the self-report measures; Table B2 for Pearson correlations). Partial eta squared (η_p^2) was reported as effect size measure. Cohen's (1988) estimates of effect sizes have been found to be misleading to interpret η_p^2 (Morris & Fritz, 2013). Quartile values of .08, .18, and .41 reported by Morris and Fritz (2013) might thus serve as guidelines to interpret η_p^2 [26]. For a better understanding and later interpretation of the results, post-hoc analyses of statistical power were calculated by participants and by items (see Table B27).

Both by-participants data, including questionnaire data, and by-items data were checked for outliers. None of the participants showed any unexpected fixation or response patterns. Likewise, none of the items evoked unexpected fixation patterns. Requirements for statistical analyses were proofed to be met before inferential analyses were done.

---

[24] In psycholinguistics, by-participants and by-items results are commonly reported one after another in one single section. This is against a common practice in psychology and many other disciplines to report the results of multiple analyses separately. Therefore, following standards in psychology, by-participants and by-items analyses and their results were reported separately. This allowed to report the results in a comprehensible way for readers with a psychologic or other non-psycholinguistic background and to avoid the impression the results obtained by participants and by items would be based on a single analysis.

[25] In by-participants analyses, the covariates could have been included when the effects of adverb gender-stereotypicality and speaker voice were tested in the context of Hypothesis 1 and Hypothesis 2. However, because the effects of the covariates could only be investigated by participants, a separate by-participants analysis was performed to test the effects of the covariates in the context of Hypothesis 3 in order to keep analyses on Hypothesis 1 and Hypothesis 2 by participants and by items comparable.

[26] The context of an experiment should be taken into account when interpreting effect sizes (Fritz et al.,2012; Morris & Fritz, 2013; Vacha-Haase & Thompson, 2004). Therefore, it needs to be considered that values reported by Morris and Fritz (2013) referred to effect sizes found in memory research. They thus rather serve as guidelines for the interpretation of effect sizes found in Experiment 1. Nonetheless, they are helpful because literature on effect size distributions in various contexts is still rare.

5.2.1 Experiment 1 – The Effects of Adverb Gender-Stereotypicality and Speaker Voice on Log-Ratios by Participants (F_1)

To test Hypothesis 1 and Hypothesis 2 by participants (F_1), a repeated measures MANOVA was conducted with log-ratios calculated by participants as a function of adverb gender-stereotypicality as a within-participants factor and speaker voice as a between-participants factor.

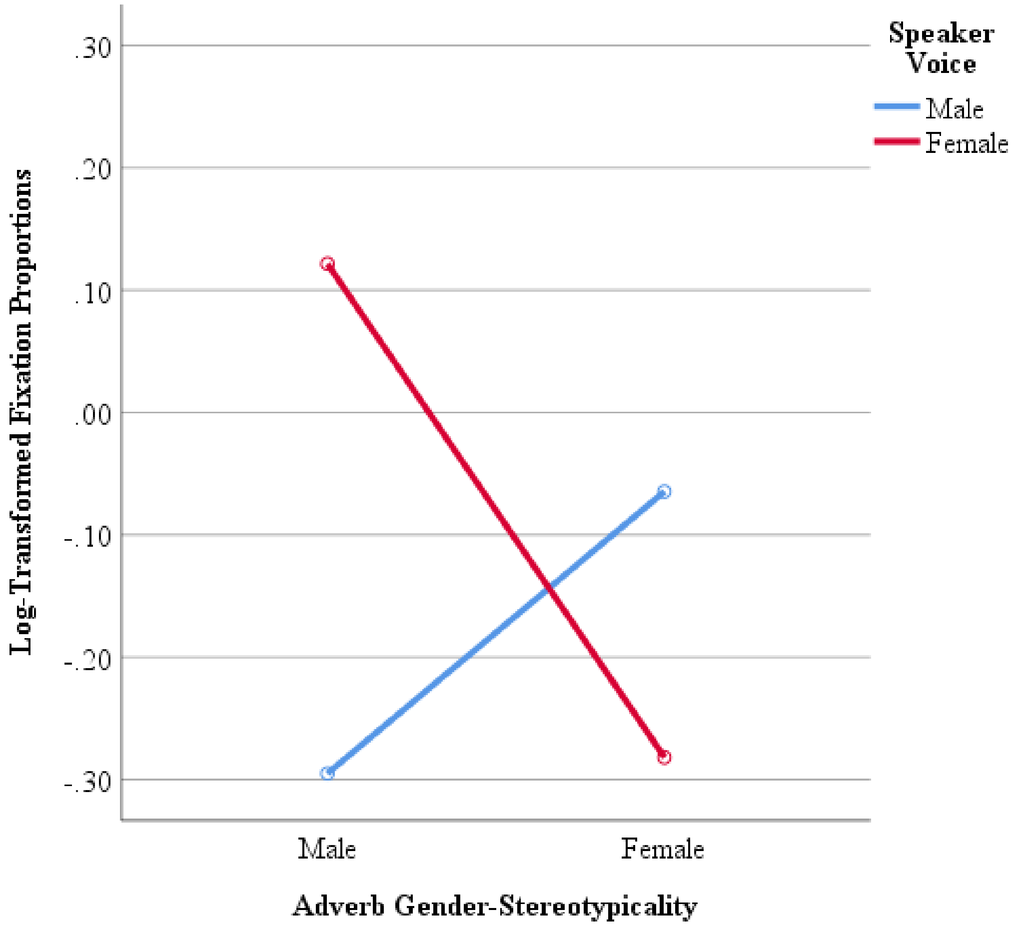
Results of the repeated measures MANOVA revealed no statistically significant main effect of adverb gender-stereotypicality on participants' log-ratios, sphericity assumed[27]: $F_1(1,83) = 0.95$, $p = .332$, $\eta_p^2 = .011$. Thus, contrary to Hypothesis 1, in analyses conducted by participants, adverb gender-stereotypicality did not guide participants' visual attention to the character whose gender matched the adverbs' gender-stereotypicality. Furthermore, the main effect of speaker voice was not statistically significant, $F_1(1,83) = 1.08$, $p = .302$, $\eta_p^2 = .013$. That is, contrary to Hypothesis 2, the speaker voice did not guide participants' eye movements to a character whose gender matched the adverbs' gender-stereotypicality, either. However, the interaction between adverb gender-stereotypicality and speaker voice was statistically significant, sphericity assumed: $F_1(1,83) = 12.80$, $p = .001$, $\eta_p^2 = .134$.

Figure 5.4 visualizes participants' log-ratios as a function of adverb gender-stereotypicality and speaker voice: Reflecting the interaction between adverb gender-stereotypicality and speaker voice, participants preferentially fixated the character whose gender matched (vs. mismatched) the adverbs' gender-stereotypicality when listening to a female speaker. Reversely, when listening to a male speaker, the stereotype-inconsistent character was more likely fixated than the gender-matching one. Overall, however, log-ratios were negative, except when stereotypically male adverbs were uttered by a female speaker. This indicates that the female character was generally more likely fixated than the male one. Only when stereotypically male adverbs were read by a female speaker, the male character tended to be more frequently looked at than the female one.

[27] As two-level within-factors were considered in the repeated measures MANOVA, the assumption of sphericity was always met (see Field, 2013).

Figure 5.4

Log-Transformed Fixation Proportions as a Function of Adverb Gender-Stereotypicality and Speaker Voice by Participants (F_1)



5.2.2 Experiment 1 – The Effects of Adverb Gender-Stereotypicality and Speaker Voice on Log-Ratios by Items (F_2)

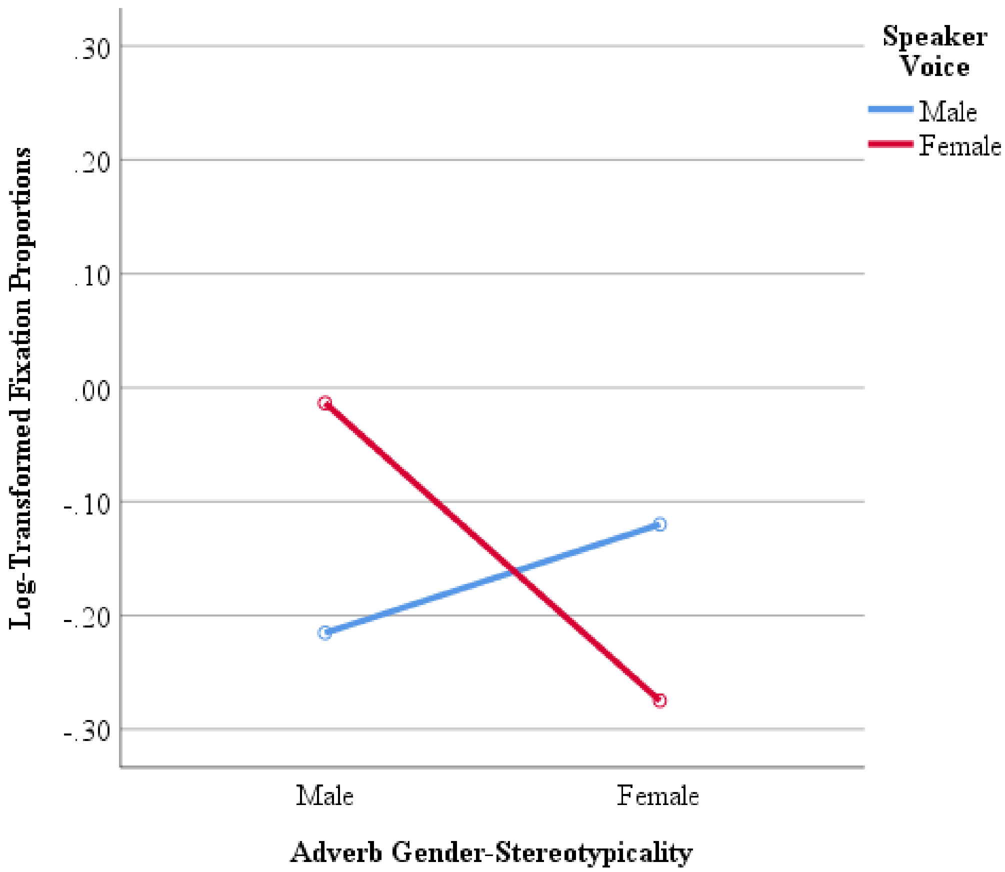
To test Hypothesis 1 and Hypothesis 2 by items (F_2), a repeated measures MANOVA was computed with log-ratios calculated by items as a function of adverb gender-stereotypicality as a between-items factor and speaker voice as a within-items factor.

In line with results obtained in by-participants analyses, the main effects of adverb gender-stereotypicality, $F_2(1,62) = 1.05$, $p = .310$, $\eta_p^2 = .017$, and speaker voice, sphericity assumed: $F_2(1,62) = 0.80$, $p = .778$, $\eta_p^2 = .001$, were not statistically significant. That is, contrary to Hypothesis 1 and Hypothesis 2, in analyses conducted by items, adverb gender-stereotypicality and speaker voice did not guide participants' eye gazes to the character whose gender matched the adverbs' gender-stereotypicality. The statistically significant interaction between adverb gender-stereotypicality and speaker voice that was found in by-participants analyses was confirmed in by-items analyses, sphericity assumed: $F_2(1,62) = 4.62$, $p = .036$, $\eta_p^2 = .069$.

Figure 5.5 illustrates log-ratios by items as a function of adverb gender-stereotypicality and speaker voice. Fixation patterns found by items resemble those found by participants: Adverbs uttered by a female speaker evoked more fixations on the character whose gender matched (vs. mismatched) the adverbs' gender-stereotypicality. Adverbs read by a male speaker evoked more looks at the stereotype-inconsistent (vs. gender-matching) character. However, log-ratios were negative irrespective of adverb gender-stereotypicality and speaker voice. This indicates that items more likely evoked fixations on the female (vs. male) character, irrespective of adverb gender-stereotypicality and speaker voice.

Figure 5.5

Log-Transformed Fixation Proportions as a Function of Adverb Gender-Stereotypicality and Speaker Voice by Items (F_2)



5.2.3 Experiment 1 – The Effects of Adverb Gender-Stereotypicality, Speaker Voice, and Participants' Gender-Related Attitudes on Log-Ratios

To test Hypothesis 3a and Hypothesis 3b, a repeated measures MANCOVA was conducted with data processed by participants (F_1). Log-ratios calculated by participants were investigated as a function of adverb gender-stereotypicality as a within-participants factor and speaker voice as a between-participants factor. Participants' endorsement of benevolent and hostile sexism, normative gender role orientation, internal and external motivation to control for sexist responses, and social desirability were considered as covariates.

In line with by-participants analyses in which the covariates were not considered, the main effects of adverb gender-stereotypicality, sphericity assumed: $F_1(1,77) = 2.63$, $p = .109$, $\eta_p^2 = .033$, and speaker voice, $F_1(1,77) = 0.76$, $p = .385$, $\eta_p^2 = .010$, were not statistically significant. The interaction between adverb gender-stereotypicality and speaker voice, however, was statistically significant, sphericity assumed: $F_1(1,77) = 10.83$, $p = .002$, $\eta_p^2 = .123$.

Regarding the main effects of the covariates, only the main effect of participants' endorsement of benevolent sexism was statistically significant. The main effects of hostile sexism, normative gender role orientation, internal and external motivation to control for sexist responses, and social desirability on participants' eye gazes were not statistically significant (see Table 5.2).

Table 5.2

Main Effects of the Covariates on Participants' Log-Ratios Within the Adverb Region

Covariate	F	p	η_p^2
Constant	< 0.01	.980	< .001
Benevolent Sexism	4.18	.044	.051
Hostile Sexism	3.79	.055	.047
Internal MCSR	0.78	.381	.010
NGRO	0.65	.423	.008
External MCSR	0.09	.760	.001
Social Desirability	0.01	.920	< .001

Note. *df*(1,77) for all main effects. NGRO = Normative Gender Role Orientation, MCSR = Motivation to Control for Sexist Responses.

To investigate the direction of the main effect of benevolent sexism on participants' eye gazes in more detail, Pearson correlations were computed. The more benevolent sexist attitudes participants reported, the more likely they fixated the male character when listening to stereotypically male adverbs, $r(84) = .19$, $p = .074$, and when listening to stereotypically female adverbs $r(83) = .18$, $p = .092$.

Regarding the interaction effects between adverb gender-stereotypicality and the covariates, there was a statistically significant interaction effect between adverb gender-stereotypicality and hostile sexism on participants' eye gazes (see Table 5.3).

Table 5.3

Interaction Effects Between Adverb Gender-Stereotypicality and the Covariates

Adverb Gender-Stereotypicality x Covariate	F	p	η_p^2
… Hostile Sexism	4.39	.039	.054
… External MCSR	2.99	.088	.037
… Social Desirability	1.56	.215	.020
… Benevolent Sexism	0.56	.456	.007
... Internal MCSR	0.52	.472	.007
… NGRO	0.33	.566	.004

Note. df(1,77) for all interaction effects. NGRO = Normative Gender Role Orientation, MCSR = Motivation to Control for Sexist Responses.

Pearson correlations were performed to investigate the direction of the interaction effect between adverb gender-stereotypicality and hostile sexism in more detail. The higher participants' hostile sexism, the more likely they fixated the female character when listening to stereotypically male adverbs, $r(84) = -.07$, p = .536, and the male character when listening to stereotypically female adverbs, $r(83) = .13$, $p = .242$ (see also Appendix B, Table A1 for participants' mean scores; see Table B2 for the full set of Pearson correlations).

Following the principle of parsimony (Tabachick & Fidell, 2007), all covariates except for participants' benevolent and hostile sexism were removed in a further step and the MANCOVA was run again. This was done to test whether the statistically significant main effect of benevolent sexism and the statistically significant interaction effects between adverb gender-stereotypicality and speaker voice and adverb gender-stereotypicality and hostile sexism would be confirmed in a more parsimonious model. The results were mainly the same as in the analysis in which all covariates were included. The statistically significant interaction between adverb gender-stereotypicality and speaker voice, sphericity assumed: $F_1(1,81) = 12.41$, $p = 001$, $\eta_p^2 = .133$, and the statistically significant main effect of benevolent sexism, $F_1(1,81) = 7.97$, $p = .006$, $\eta_p^2 = .090$, were confirmed. The interaction between adverb gender-stereotypicality and hostile sexism was however not statistically significant, sphericity assumed: $F_1(1,81) = 2.99$, $p = .087$, $\eta_p^2 = .036$.

Summing up, when the covariates were considered, the interaction between the adverbs' gender-stereotypicality and speaker voice and the main effect of benevolent sexism determined whether the male or the female character was looked at. Contrary to Hypothesis 3a, however, benevolent sexism guided participants' eye gazes to the male character, independent of the adverbs' gender-stereotypicality. That is, Hypothesis 3a predicting the higher participants' levels of benevolent and hostile sexism and normative gender-role orientation, the more they would look at the character whose gender matched (vs. mismatched) the adverbs' gender-stereotypicality, was not confirmed. The same was true for Hypothesis 3b predicting the higher participants' levels of internal and external motivation to control for sexist responses and social desirability, the less they would fixate the character whose gender matched (vs. mismatched) the adverbs' gender-stereotypicality.

5.3 Experiment 1 – Manipulation Check

Inferential analyses of the NP2 region served as a manipulation check to confirm the impression conveyed by the time course graphs that the NP2 target was looked when it was explicitly mentioned at the end of the sentence (see Figure 5.2 and Figure 5.3). This was crucial because fixations on the NP2 target would indicate that participants could link the sentences' content to the displayed characters. Therefore, it was important that the NP2 target was looked at, regardless of whether participants had been exposed to a male or a female speaker. To analyze fixation patterns at the end of the sentence, repeated measures MANOVAs were performed with log-ratios by participants (F_1) and by items (F_2) within the NP2 region as a function of NP2 gender and speaker voice.

5.3.1 Experiment 1 – The Effects of NP2 Gender and Speaker Voice on Log-Ratios by Participants (F_1)

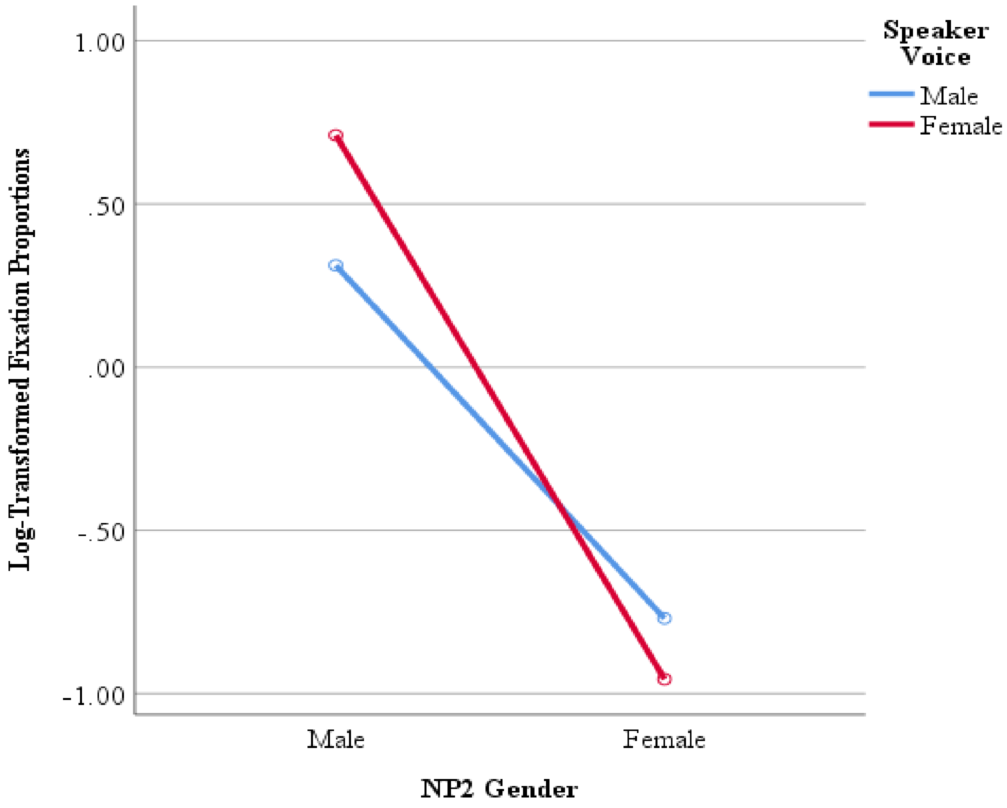
A repeated measures MANOVA was performed with log-ratios within the NP2 region calculated by participants (F_1) as a function of NP2 gender as a within-participants factor and speaker voice as a between-participants factor.

The main effect of NP2 gender was statistically significant, sphericity assumed: $F_1(1,83) = 234.64$, $p < .001$, $\eta_p^2 = .739$. That is, confirming the time course graphs, the NP2 target was looked at as soon as it was mentioned at the end of the sentence. The main effect of the speaker voice was not statistically significant, $F_1(1,83) = 2.29$, $p = .134$, $\eta_p^2 = .027$. This indicates that the NP2 target was looked at regardless of whether it was uttered by a male or a female speaker. However, unexpectedly, the interaction between NP2 gender and speaker voice was statistically significant, sphericity assumed: $F_1(1,83) = 10.63$, $p = .002$, $\eta_p^2 = .114$.

Figure 5.6 visualizes participants' fixation patterns within the NP2 region as a function of NP2 gender and speaker voice: The NP2 target was looked at when it was named, irrespective of the speaker voice. However, it was more likely looked at when it was uttered by a female (vs. male) speaker voice.

Figure 5.6

Log-Transformed Fixation Proportions as a Function of NP2 Gender and Speaker Voice by Participants (F_1)



5.3.2 Experiment 1 – The Effects of NP2 Gender and Speaker Voice on Log-Ratios by Items (F_2)

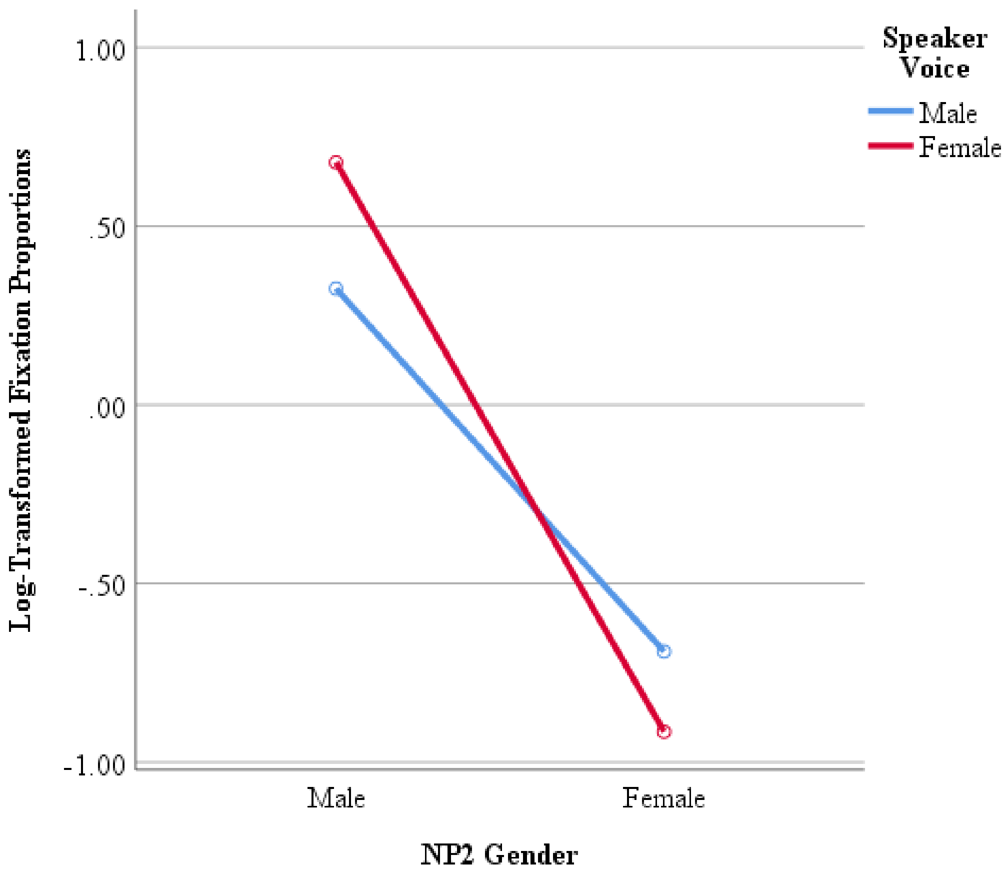
A repeated measures MANOVA was conducted with log-ratios calculated by items as a function of NP2 gender as a between-items factor and speaker voice as a within-items factor.

In line with by-participants findings, the main effect of NP2 gender was statistically significant in analyses by items, $F_2(1,62) = 285.00$, $p < .001$, $\eta_p^2 = .821$, while the main effect of speaker voice was not statistically significant, sphericity assumed: $F_2(1,62) = 1.03$, $p = .314$, $\eta_p^2 = .016$. The interaction between NP2 gender and speaker voice was statistically significant, sphericity assumed: $F_2(1,62) = 20.76$, $p < .001$, $\eta_p^2 = .251$.

Figure 5.7 illustrates fixation patterns by items: In line with fixation patterns by participants, the NP2 target was looked at when it was mentioned, regardless of the speaker voice. However, it was more likely fixated when it was uttered by a female (vs. male) speaker.

Figure 5.7

Log-Transformed Fixation Proportions as a Function of NP2 Gender and Speaker Voice by Items (F_2)



5.4 Experiment 1 – Exploratory Analyses

Exploratory analyses were conducted to explain the results in more detail (see Appendix B, pp. 299).

6 Experiment 1 (Adverbs – Humans) – Discussion

Experiment 1 was the first of four visual world experiments (see Cooper, 1974; Magnuson et al., 1998; Tanenhaus et al., 1995). In Experiment 1, I focused on three main research aspects: First, I investigated whether gender-stereotypical adverbs would guide participants' visual attention to a target whose gender matched (vs. mismatched) their gender-stereotypicality (Hypothesis 1). This was innovative in that sense that although certain attributes are stereotypically ascribed to men and women (e.g., Abele, 2003; Abele & Wojciszke, 2007; Bem, 1974; 1981; Berger & Krahé, 2013), their impact on language comprehension had been widely neglected in prior psycholinguistic research. Second, the effects of the social context in terms of the speaker voice were taken into account (see Münster & Knoeferle, 2018). That is, the effect of the adverbs' gender-stereotypicality on participants' anticipatory eye movements was hypothesized to be enhanced when the adverbs would be read by a speaker whose gender matched (vs. mismatched) their gender-stereotypicality (Hypothesis 2). Third, I explored whether participants' benevolent and hostile sexism, and normative gender role orientation guided their eye gazes to the gender-matching character (Hypothesis 3a). Reversely, participants' internal and external motivation to control for sexist responses and social desirability were expected to lead to less looks at the gender-matching character. This served to answer researchers' call to view language processing as a social phenomenon that is affected by the context in which language is encountered and by participants' individual characteristics, such as their attitudes and beliefs (Münster & Knoeferle, 2018; Van Berkum et al., 2008; 2009).

Contrary to Hypothesis 1, that the adverbs' gender-stereotypicality would guide participants' visual attention, the main effect of adverb gender-stereotypicality was not statistically significant. That is, participants did not look at the character whose gender matched the adverbs' gender-stereotypicality. Contrary to Hypothesis 2, the main effect of speaker voice was not statistically significant, either. That means, participants' visual attention was not directed at the gender-matching character when the speaker's gender matched the adverbs' gender-stereotypicality. The interaction between adverb gender-stereotypicality and speaker voice was statistically significant: The character whose gender matched (vs. mismatched) the adverbs' gender-stereotypicality was more likely fixated when the adverbs were articulated by a female speaker. When the adverbs were uttered by a male speaker, the character that mismatched (vs. matched) the adverbs' gender-stereotypicality, the stereotype-inconsistent character, was more likely fixated. This fixation pattern was found in analyses conducted by participants and by items.

Though Hypothesis 1 was not confirmed, this fixation pattern suggests that participants could seemingly differentiate between stereotypically male and female adverbs. This differentiation might have been facilitated by the fact that the adverbs were taken from a set of attributes that was judged as gender-stereotypical according to Western Society's beliefs about men and women. It is thus very likely that participants have been repeatedly confronted with them in their life. This may have strengthened associations between the adverbs and male and female gender which may have been activated automatically as soon as the adverbs were encountered (see e.g., Devine, 1989; Fazio, 1990; 2007; Fazio et al., 1995; see also Bargh, 1999 for a review). That way, participants may have been able to easily link the adverbs to the characters and thus to grasp the meaning of the sentence. This would be in line with the Social CIA (Münster & Knoeferle, 2018) which postulates that as soon as verbal information is encountered, participants draw an initial link between verbal and visual information. This link is reconciled with ongoing information, context factors, participants' knowledge, demographics, and beliefs. If necessary, this link is revised, which is then reflected by shifts of visual attention. Interpreted in the context of the Social CIA, the finding that participants fixated the stereotype-inconsistent character when listening to a male speaker voice might indicate that the speaker voice was used to revise the initially established link between the adverbs and either of the characters. However, the Social CIA assumes that an initial link is revised to predict the sentence's target and thus to gain language comprehension. Fixating the stereotype-inconsistent character when listening to a male speaker voice seemed not beneficial in terms of language comprehension because, following gender stereotypes, it was unlikely that the stereotype-inconsistent character was the sentence's target (see e.g., Duffy & Keir, 2004; Pyykkönen et al., 2010). Extending the Social CIA and most prior psycholinguistic research (see e.g., Allopenna et al., 1998; Cooper, 1974; Knoeferle & Crocker, 2006; 2007; Tanenhaus et al., 1995), participants' fixation pattern might therefore go beyond the reflection of their attempts to predict incoming verbal information. It is possible that, when listening to a male speaker, participants have countered their associations between the adverbs and the characters by avoiding fixations on the character whose gender matched the adverbs' gender-stereotypicality. Because fixations on the two characters were interrelated (see Arai et al., 2007), this may have led to more fixations on the stereotype-inconsistent character. Such counteractive or controlled responses are possible when participants are aware of stereotype content, have cognitive capacities, and are motivated to do so (see Conrey et al.,2005; Devine, 1989; Fazio, 1990; 2007). Indicating their motivation, participants reported a relatively high internal amongst a low external motivation to control for sexist responses. Assuming that the adverbs have activated gender stereotypes automatically and that thus language comprehension was accomplished easily, participants may have had cognitive capacities to be aware of gender stereotypes and to counter them by avoiding fixations on the gender-matching character accordingly (see Schneider & Chein, 2003; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977 for automatic vs. controlled processes). To illustrate, none of the participants guessed the actual purpose of the experiment. Only a few conjectured the experiment might be about men's and women's professional roles. It is therefore unlikely that participants were fully aware of the adverbs' gender-stereotypicality. However, after some trials, they may have noticed that the sentence's target was always depicted on the screen amongst its male or female counterpart. Moreover, as the speaker voice varied between participants, participants may have noticed whether the speaker voice was male or female after some trials. Due to this constellation, participants may have sensed that the sentences could somehow be related to gender stereotypes. Participants' relatively low levels of sexism and relatively high levels of internal motivation to control for sexist responses further indicate that participants might have internalized the aim to behave non-sexist. Therefore, the sense to encounter gender stereotypes likely caused fear and discomfort to possibly have behaved in a gender-stereotypical or sexist manner. To resolve this discomfort, participants might have reflected their past fixation behavior. At the same time, their vigilance for indications of gender stereotypes and sexism might have been enhanced (see Monteith, 1993; Monteith et al.,2002). It thus appears plausible that above and beyond the interaction between adverb gender-stereotypicality and speaker voice, participants' sexist attitudes and participant gender guided their visual attention.

To illustrate, testing Hypothesis 3a and Hypothesis 3b, amongst the interaction between adverb gender-stereotypicality and speaker voice, the main effect of benevolent sexism and the interaction effect between adverb gender-stereotypicality and hostile sexism were statistically significant when the effects of all the covariates (i.e., benevolent and hostile sexism, normative gender role orientation, internal and external motivation to control for sexist responses, and social desirability) were considered. However, only the interaction between adverb gender-stereotypicality and speaker voice and the main effect of benevolent sexism were confirmed in a more parsimonious model in which only benevolent and hostile sexism was considered as a covariate. Pearson correlations were performed to investigate the direction of this effect in more detail. Contrary to Hypothesis 3a, predicting participants' benevolent and hostile sexism and normative gender role orientation would lead to more looks at the gender-matching character, the higher participants' endorsement of benevolent sexism, the more they looked at the male character, independent of the adverbs' gender-stereotypicality. Contrary to

Hypothesis 3b, participants' internal and external motivation to control for sexist responses and social desirability did not guide their eye gazes when listening to the adverbs.

To explain the main effect of benevolent sexism, perhaps due to benevolent sexists' view of men taking action (see Glick & Fiske, 2001), the more benevolent sexist attitudes participants indicated, the more they may have anticipated the male character to be the sentence's target. Due to a subjectively positive view of women which also goes along with benevolent sexism (see Glick & Fiske, 2001), the more participants endorsed of benevolent sexism, the less they might have regarded benevolent sexist attitudes as a form of prejudice (see Barreto & Ellemers, 2005a; Glick & Fiske, 2001). In turn, the less they might have controlled or inhibited anticipatory eye movements to the male character.

Moreover, in exploratory analyses, the main effect of participant gender was statistically significant. Compared to female participants, male participants more likely fixated the character whose gender matched the adverbs' gender-stereotypicality when listening to a female speaker and the stereotype-inconsistent character when listening to a male speaker.

To better understand these findings, it needs first to be clarified why a male speaker voice seemed to have led to counteractive fixations while a female voice did apparently not. According to Western Society, a male speaker represented a member of the dominant group from which gender stereotypes and sexism emanate (e.g., Baron et al.,1991; Barreto & Ellemers, 2005b; Calvert & Ramsey, 1996; Fiske et al.,1999; Glick & Fiske, 1996) and who benefits from their persistence (see Glick & Fiske, 1996; 2001; Sidanius & Pratto, 1999; 2012). As such, prior research has demonstrated that sexist content was detected more easily and perceived as more intense when expressed by a male (vs. female) agent (Baron et al., 1991; Barreto & Ellemers, 2005b). Therefore, when listening to a male speaker, participants' vigilance for gender stereotypes and their discomfort to possibly behave gender-stereotypical or sexist may have been enhanced compared to when listening to a female speaker voice. As such, participants might have avoided to look at the gender-matching character when listening to the male speaker which then led to fixations on the stereotype-inconsistent character. Assuming that the aim to respond non-stereotypical and non-sexist was internalized, such counteractive fixations may have been performed automatically (see De Houwer, 2009). Maybe, as members of the dominant group themselves, male participants felt particular discomfort and fear to possibly behave in a sexist manner. This may then have affected their fixations in that sense that the interaction between the adverbs' gender-stereotypicality and speaker voice more strongly affected their gazes than female participants' gazes (see Lowery et al.,2001 for similar results in the context of ethnicity). Remarkably, the role of participant gender and of speaker voice was also emphasized by further exploratory analyses: The interactions between participant gender and speaker voice had statistically significantly affected participants' responses on benevolent and hostile sexism and normative gender role orientation: Participants indicated lower levels of benevolent and hostile sexism after having been exposed to a speaker of their opposite gender. At a descriptive level, female participants indicated particularly low levels of benevolent sexism after having been exposed to a male speaker. Male participants indicated particularly low levels of hostile sexism and normative gender role orientation after having listened to a female speaker.

These response patterns suggest that the speaker voice and participants' own group membership in terms of gender might have been processed on a deep level which is possible only if participants had paid close attention to both over time (see Baddeley, 2000; Craik & Lockhart, 1972). This in turn supports the conjecture that participants' vigilance for factors that might be related to gender stereotypes may have been enhanced during the experiment.

Male and female response patterns on the self-report measures further suggest that social norms might have played a role during the experiment. Perhaps female participants had a particular interest not to appear as if taking advantage of benevolent sexists' subjectively positive view of women and the prosocial behavior it might elicit (Glick & Fiske, 2001) after having listened to a male speaker. Likewise, it is possible that male participants had a particular interest not to express agreement to blatant sexist views of women and to traditional gender roles after having listened to a female speaker because this would contrast non-sexist and egalitarian views of men and women. The assumption that participants might have attempted to follow social norms is further indicated by the fact that, amongst low levels of sexism and a high internal motivation to control for sexist responses, participants reported relatively high levels of social desirability. Student samples were found to indicate relatively low levels of sexism (De Judicibus & McCabe, 2001; Pettijohn & Walzer, 2008). Therefore, behaving non-sexist is possibly regarded a norm among students which may be linked to the fear of disapproval for not following it. Perhaps participants did not strive to respond in a non-stereotypical and non-sexist manner solely because they truly aimed to be non-sexist, but also because they aimed to identify with their peers' norms and values (see Boninger et al.,1995). Amongst the auditory presence of a male or female speaker, their peers' norm to behave non-sexist may have been activated by the experimenter, a student peer, being around during the experiment taking place on a university campus.

Participants' attempts not to behave in a gender-stereotypical or sexist manner seemed to have led them to reflect a particularly positive view of women (see also Eagly &

Mladinic, 1989). This was indicated by exploratory analyses examining the effect of adverb connotation. Though the interaction between adverb connotation and speaker voice was statistically significant only in analyses conducted by participants, by-participants and by-items analyses revealed the same fixation pattern at a descriptive level: Independent of adverb gender-stereotypicality, adverb connotation, and speaker voice, the female character was preferred over the male one. The preference for the female character was most apparent when positively connoted adverbs were uttered by a male speaker and when positively connoted adverbs were stereotypically female. Guerra and colleagues (2021) also found a preference for a female (vs. male) character. They conjectured that this preference may reflect participants' sexist attitudes. The results of the present Experiment 1 might complement their findings. It is possible that not participants' sexist attitudes per se led to the preference for the female character, but their attempts to counteract gender-stereotypes and sexist attitudes. This assumption might be strengthened by the finding that, though not statistically significant in a more parsimonious model, participants were more likely to look at the stereotype-inconsistent character, the higher their levels of hostile sexism.

Independent of the speaker voice, the sentence's target was looked at as soon as it was named explicitly at the end of the sentence. This was indicated by a statistically significant main effect of the NP2 gender and a statistically non-significant main effect of speaker voice found in analyses conducted by participants and by items. Time course graphs showed that a female NP2 target was fixated even shortly before its onset. Corrective eye movements from the character whose gender matched (vs. mismatched) the adverbs' gender-stereotypicality to the one that matched the NP2 target's actual gender were performed if necessary. Again, in line with the Social CIA (Münster & Knoeferle, 2018; see also Knoeferle & Crocker, 2006; 2007), this implies that language processing is an incremental process in which predictions about language are revised as long as verbal information is ongoing. Moreover, these results showed that, if available, participants rely on explicit verbal information rather than on stereotypes to gain language comprehension (Knoeferle & Crocker, 2006; 2007; see also e.g., Duffy & Keir, 2004; Irmen, 2007; Pyykkönen et al., 2010; Rodriguez et al., 2016). However, a statistically significant interaction between the NP2 gender and speaker voice demonstrated that the correct target was more likely to attract participants' visual attention when it was specified by a female (vs. male) speaker. Recall that within the adverb region, participants who had listened to a male speaker were more likely to counter the adverbs' gender-stereotypicality. Maybe because of that, they needed longer to comprehend that now one of the characters was explicitly referred to. Moreover, according to the time course graphs, participants seemed to more likely look at the NP2 target when it was female (vs. male). This tendency was also found in Guerra, et al. (2021). It may have been due to the target role's suffix. More precisely, in German, the suffix "-in" clearly marks a target as female, while the male form is often generically used for male and female targets (see Irmen, 2007; Irmen & Roßberg, 2004). Thus, when listening to a role name's male form, both characters were likely to be the correct NP2 target. Listening to the female form, only the female character could be the correct target.

To sum, the results from Experiment 1 revealed two important insights: First, though the hypotheses were not confirmed, the adverbs' gender-stereotypicality seemed to have automatically activated stereotypical associations with men and women. Second, participants' associations determined their language processing. That way, in line with the Social CIA (Münster & Knoeferle, 2018), language processing was shown to be an incremental process that seemed to be affected by the interplay between verbal and visual information, participant gender, participants' stereotypes and sexist attitudes, and the social context in terms of the speaker voice and possibly also in terms of the experimental setting of a university campus. Extending the Social CIA and most prior linguistic literature, language processing apparently went beyond the acquisition of language comprehension. It seems that, when listening to gender-stereotypical adverbs, participants attempted to counter gender stereotypes and sexism according to their attitudes and apparently also in line with prevailing social norms. To do so, they seemingly countered fixations on the gender-matching character, particularly when listening to a male speaker voice. To be aware of gender stereotypes and to counter them accordingly, may have been facilitated by the fact that language comprehension was obtained effortlessly because the adverbs seemed to be automatically linked to either of the characters.

7 Experiment 2 (Main Verbs - Humans) – Hypotheses

The hypotheses were analogous to Experiment 1, except that the main verbs' gender-stereotypicality was in focus of investigation according to prior research (e.g., Guerra et al., 2021; Rodriguez et al., 2015; 2016):

1 The main verbs' gender-stereotypicality and

2 a speaker voice whose gender matches (vs. mismatches) the main verbs' gender-stereotypicality

lead to higher log-gaze probabilities to look at a character whose gender matches (vs. mismatches) the main verbs' gender-stereotypicality[1]

Log-gaze probabilities to look at a character whose gender matches (vs. mismatches) the main verbs' gender-stereotypicality are <u>higher</u>, the higher participants' values on

3a Benevolent and hostile sexism and normative gender role orientation

and <u>lower</u>, the higher participants' values on

3b Internal and external motivation to control for sexist responses and social desirability.

8 Experiment 2 (Main Verbs - Humans) – Method

8.1 Pretest IV – Evaluation of Noun and Verb Phrases' Gender-Stereotypicality

Pretest IV was done to extent the existing item-pool of grammatically and stereotypically gender-neutral noun phrases and gender-stereotypical verb phrases because more noun-verb combinations were needed to create useful sentences for Experiment 2.

8.1.1 Pretest IV – Procedure and Noun and Verb Phrases' Suitability

Procedure and implementation of Pretest IV were similar to Pretest II (see Section 4.2). Participants were given a set of 113 noun and verb phrases each whose gender-stereotypicality they had to judge independently of one another. Some noun and verb phrases were taken from Pretest II, others were newly added. To find new noun-verb combinations suitable for Experiment 2, noun and verb phrases were presented multiple times, but they were always anteceded or followed by another noun or verb phrase, respectively. Because the verb phrase would directly follow the noun phrase in the experimental sentences of Experiment 2, it was particularly important to rule out confounding effects of the anteceding noun phrase's gender-stereotypicality as far as possible. Therefore, a noun phrase (NP1) had to be judged as stereotypically gender-neutral irrespective of the following verb phrase. Reversely, it was crucial that a verb phrase was judged as gender-stereotypical irrespective of the anteceding noun phrase. That is, to be suitable for the experimental sentences, a noun phrase had to be judged as stereotypically gender-neutral every time it was presented in Pretest IV. Likewise, participants' evaluations of a verb phrase's gender-stereotypicality had always to be the same irrespective of an anteceding noun phrase.

8.1.2 Pretest IV – Sample

N = 39 participants[4] completed Pretest IV (male: n = 6, female: 33; M_{age} = 26.36, SD_{age} = 6.52, age range: 18-52 years). Most of them were students (students: n = 34, professionals: n = 5) of German nationality (German nationality: n = 37, other nationality: n = 2). All participants had a good command of German language (German native speaker: n = 38, other native language: n = 1).

8.1.3 Pretest IV – Results

In total, 36 noun phrases were rated as stereotypically male, 14 noun phrases were considered stereotypically female, and 63 noun phrases were identified as stereotypically gender-neutral. Likewise, 43 verb phrases were evaluated as stereotypically male, while 52 verb phrases were considered stereotypically female, and 19 verb phrases were identified as stereotypically gender-neutral (see Appendix A, Table A23 – Table A28)[6].

8.2 Experiment 2 – Sample

A student sample was recruited at Bielefeld University. Eligibility requirements were the same as in Experiment 1 (see Section 4.4). $N = 85$ participants[4] took part in Experiment 2. Seven of them had to be excluded from data analyses due to technical issues with the eye tracker. Most of the remaining $n = 78$ (male: $n = 39$, female: 39; $M_{\text{age}} = 22.64$, $SD_{\text{age}} = 2.93$, age range: 18-32 years) were German nationals (German nationality: $n = 76$, other nationality: $n = 2$). All respondents were German native speakers with normal or corrected-to normal vision and audition who gave oral and written consent to have their data recorded. They ensured to have understood the experiment instructions and could correctly answer the questions followed by the filler trials. All participants were naïve about the purpose of Experiment 2.

8.3 Experiment 2 – Design

A mixed design was realized with main verb (V2) gender-stereotypicality (male vs. female) and NP2 gender (male vs. female) as within-participants factors and with speaker voice (male vs. female) as a between-participants factor.

8.4 Experiment 2 – Verbal Stimuli: The Experimental Sentences

The structure of the experimental sentences was similar to Experiment 1. In German subject-verb-object (SVO) sentences, an object was described to be manipulated by a representative of a gender-stereotypical profession, e.g., “Das Fenster wird ausgemessen von dem Bauarbeiter.” [The window is sized by the construction worker.].

32 experimental tuples were created as illustrated in Table 8.1. The sentences of a tuple were identical, except for the main verb. The object that was described to be manipulated was represented by a noun that was gender-neutral in terms of grammatical gender and gender-stereotypicality. In one sentence of a tuple, the main verb was stereotypically male, in the other, it was stereotypically female. To avoid a bias due to word length (e.g., Hyönä & Olson, 1995; Inhoff & Radach, 1998), two main verbs that had the same number of phonemes were assigned to the same tuple, if possible. The NP2 target was a representative of a gender-stereotypical profession. Stereotypically male professions were used in the male form, while stereotypically female professions were used in the female form. In half of the tuples, the NP2 target was male and female, respectively. To consider the fact that participants in Pretest IV might not have been able to judge a verb phrase’s gender-stereotypicality independent of an anteceding noun phrase, nouns and verbs were combined that had been presented one after another in Pretest IV,

if possible. At the same time, NP1, main verbs, and NP2 targets were combined in a way the sentences sounded as plausible as possible (see also Appendix A, Table A29 for the full set of experimental sentences).

To investigate the impact of the speaker voice on participants' anticipatory eye movements, each sentence was read by the same male and female speakers who had already recorded the sentences of Experiment 1. As such, the sentences were read and, at a later point, edited analogous to Experiment 1 (see Section 4.6.1).

Table 8.1

Schematic of the Experimental Sentences Within- and Between-Tuples

Tuple	NP1	VP		Prep.	NP2
1	Das Fenster	wird	**ausgemessen**	von	dem **Bauarbeiter.**
	[The window	is	**sized**	by	the **construction worker**.]
		stereotypically **male** main verb			stereotypically **male** profession
	Das Fenster	wird	**geputzt**	von	dem **Bauarbeiter.**
	[The window	is	**cleaned**	by	the **construction worker**.]
		stereotypically **female** main verb			stereotypically **male** profession
2	Das Brot	wird	**gebrochen**	von	der **Flugbegleiterin.**
	[The bread	is	**broken**	by	the **stewardess**.]
		stereotypically **male** main verb			stereotypically **female** profession
	Das Brot	wird	**gesucht**	von	der **Flugbegleiterin.**
	[The bread	is	**searched**	by	the **stewardess**.]
		stereotypically **female** main verb			stereotypically **female** profession

Note. The original version of a set of experimental sentences and their translation to English in parentheses below. NP = Noun Phrase, VP = Verb Phrase, Prep. = Preposition.

8.5 Experiment 2 – Verbal Stimuli: The Filler Sentences

40 filler sentences were added to the set of 32 experimental tuples. The fillers just served to prevent participants from guessing the hypotheses. They were not used in later statistical analyses.

The fillers in Experiment 2 were very similar to those in Experiment 1 (see Section 4.7). The object that was described as being manipulated (NP1) was represented by a grammatically gender-neutral noun that was judged as stereotypically male or female. It was followed by a stereotypically gender-neutral main verb. The NP2 target at the end of the sentence was stereotypically male or stereotypically female. It was used in the male or in the female form so that in half of the filler sentences, the NP2 target's grammatical gender mismatched its gender-stereotypicality. The filler sentences were recorded in the same reading style and by the same speakers who had read the experimental sentences of Experiment 2. However, because the fillers were not needed for further statistical analyses, no further editing was needed. Additionally, analogous to Experiment 1, questions were formulated for each filler trial to ensure that participants paid attention to the verbal and the visual stimuli (see Appendix A, Table A30).

8.6 Experiment 2 – Visual Stimuli

The visual scenes that referred to the sentences were created in the same manner in Experiment 1 (see Section 4.8; see Appendix A, pp. 264 for the visual scenes).

8.7 Experiment 2 – Questionnaire

The questionnaire was identical to Experiment 1 (see Section 4.9; see Appendix A, Table B4 for internal consistencies and means on the measured constructs per speaker voice).

8.8 Experiment 2 – Procedure

The procedure was identical to Experiment 1 (see Section 4.10).

9 Experiment 2 (Main Verbs - Humans) – Results

9.1 Experiment 2 – The Steps of Data Processing

Data processing and data analyses were done analogously to Experiment 1, but with main verb gender-stereotypicality as an experimental factor. More precisely, in a first step, to capture participants' fixations, the same set of AIs were used as in Experiment 1 (see Figure 5.1). This was possible because the same male and female characters and mostly the same objects (or at least objects of the same size) were displayed as in Experiment 1. In a second step, log-ratios of fixations on the male relative to the female character were calculated per main verb gender-stereotypicality, per speaker voice, and per NP2 target (see Section 5.1.1). In a third step, using log-ratios, time course graphs were created to allow for a rough inspection of participants' fixation patterns over the course of the entire sentence. In a fourth step, to prepare data for inferential analyses, the main verb region and the NP2 region were predefined from word onset up to the onset of the following word region. Then, the longest main verb duration (1921ms) and NP2 duration (2597ms) were identified[28]. The length of a word region was the time in which fixations were counted from each individual item's main verb onset, NP2 onset, respectively (see Section 5.1.3)[29]. In a fifths and final step of data processing, the sum of fixations each within the predefined main verb region and within the predefined NP2 region was aggregated, restructured, and log-transformed per experimental condition (i.e., main verb gender-stereotypicality, NP2 gender, speaker voice) by participants (F_1) and by items (F_2) (see Section 5.1.4).

9.1.1 Experiment 2 – Step 3 of Data Processing: Creating Time Course Graphs

Using log-ratios calculated per main verb gender-stereotypicality and NP2 gender, graphs were created to display participants' time-lined fixation patterns over the course of the entire sentence (see Figure 9.1 and Figure 9.2). To certainly consider all fixations, time course graphs include the latest NP2 offset which was at 6615ms. In order to roughly assign gaze-patterns to word regions, mean on- and offsets of the main verb and the NP2 region were inserted. Analogous to Experiment 1, mean main verb and NP2 on- and offsets were computed considering the sum of fixations over participants and items (see Table 9.1; see also Appendix A, Table A29 for all on- and offsets).

[28] The sentences of Experiment 2 were newly recorded. Therefore, NP2 on- and offsets and durations differed from Experiment 1, although the same NP2 targets were used in Experiment 1 and Experiment 2.

[29] Three tuples had to be excluded from data analyses because in combination with a respective NP1, the main verbs' gender-stereotypicality did not meet the requirements for the experimental sentences of Experiment 2 (see Section 8.4). Consequently, only 29 tuples were considered in the data analyses.

Table 9.1

Earliest, Latest, and Mean On- and Offset Times (ms) of the Main Verb and the NP2 Region

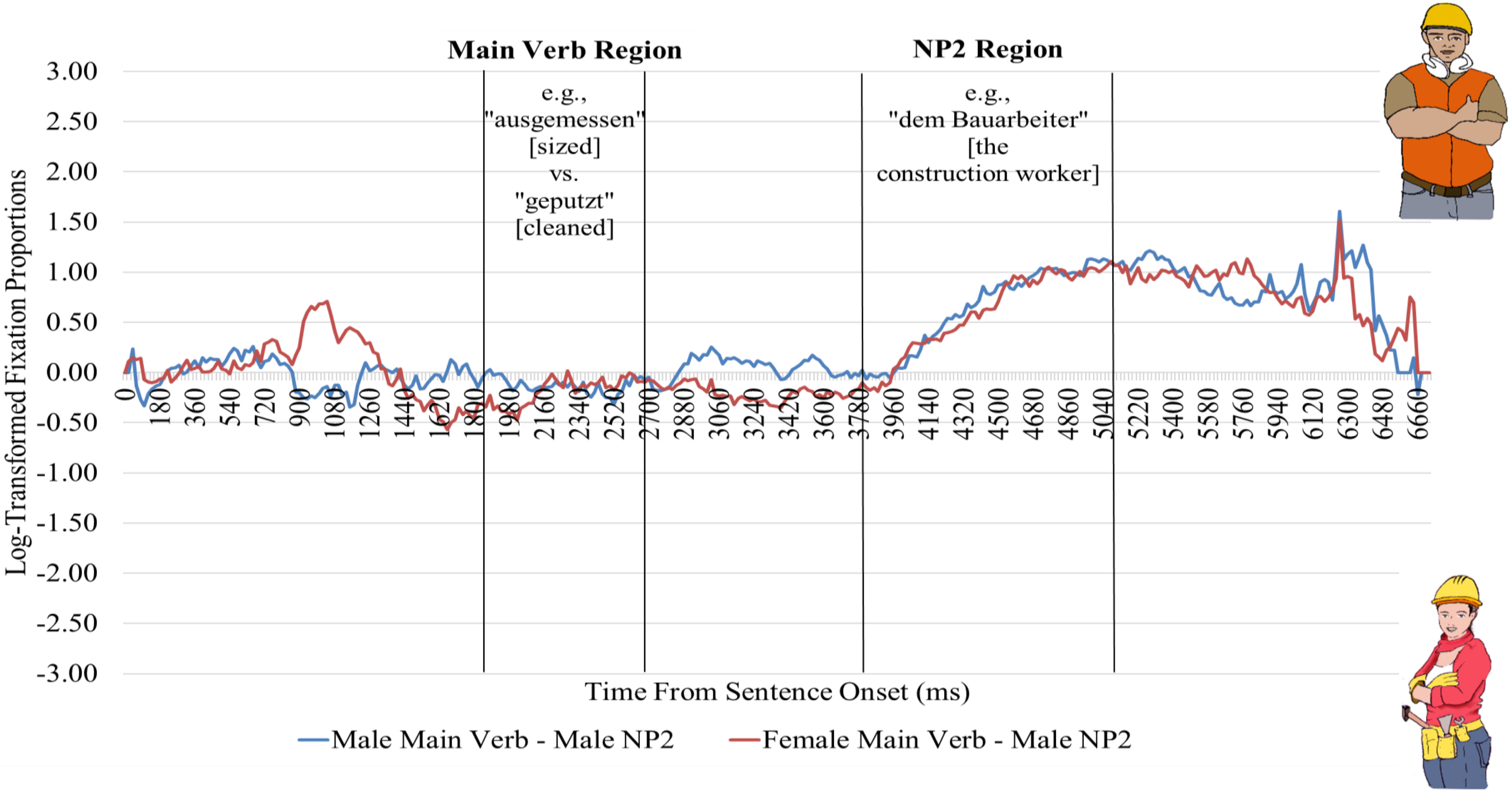
Word Region	Earliest	Latest	*M*	*SD*
Main Verb Onset	1172	2363	1852.92	224.11
Main Verb Offset	1903	3363	2685.02	281.03
NP2 Onset	3258	4440	3808.75	288.11
NP2 Offset	4252	6615	5085.52	472.15

Note. $N = 21090$ (sum of all fixations over participants and items).

Figure 9.1 visualizes time course graphs for stereotypically male and female main verbs that refer to a male NP2 target: Regarding fixations within the main verb region, from about 400ms before mean main verb onset until about 300ms after, there was a slight divergence between log-ratios for stereotypically male and female main verbs. More precisely, during this time, log-ratios for stereotypically male main verbs ranged around zero, while log-ratios for stereotypically female main verbs reached values up to about -0.50. This indicates that if any, there might have been only a slight tendency to look at the female character when listening to stereotypically female main verbs, while there was no preference for the male character when listening to stereotypically male main verbs. Shortly after mean main verb offset until shortly before mean NP2 onset, there was a slight tendency to more likely inspect the character whose gender matched the main verbs' gender-stereotypicality. During this time, log-ratios ranged between about 0.25 for stereotypically male and about -0.25 for stereotypically female main verbs. That is, if at all, there might have been a slight preference for the gender-matching character when listening to the preposition "von" [of] that refers to the NP2 target. Regarding fixation patterns within the NP2 region, the NP2 target was looked at shortly after mean NP2 onset.

Figure 9.1

Time Course Graph for Sentences With a Stereotypically Male vs. Female Main Verb Referring to a Stereotypically Male NP2 Target

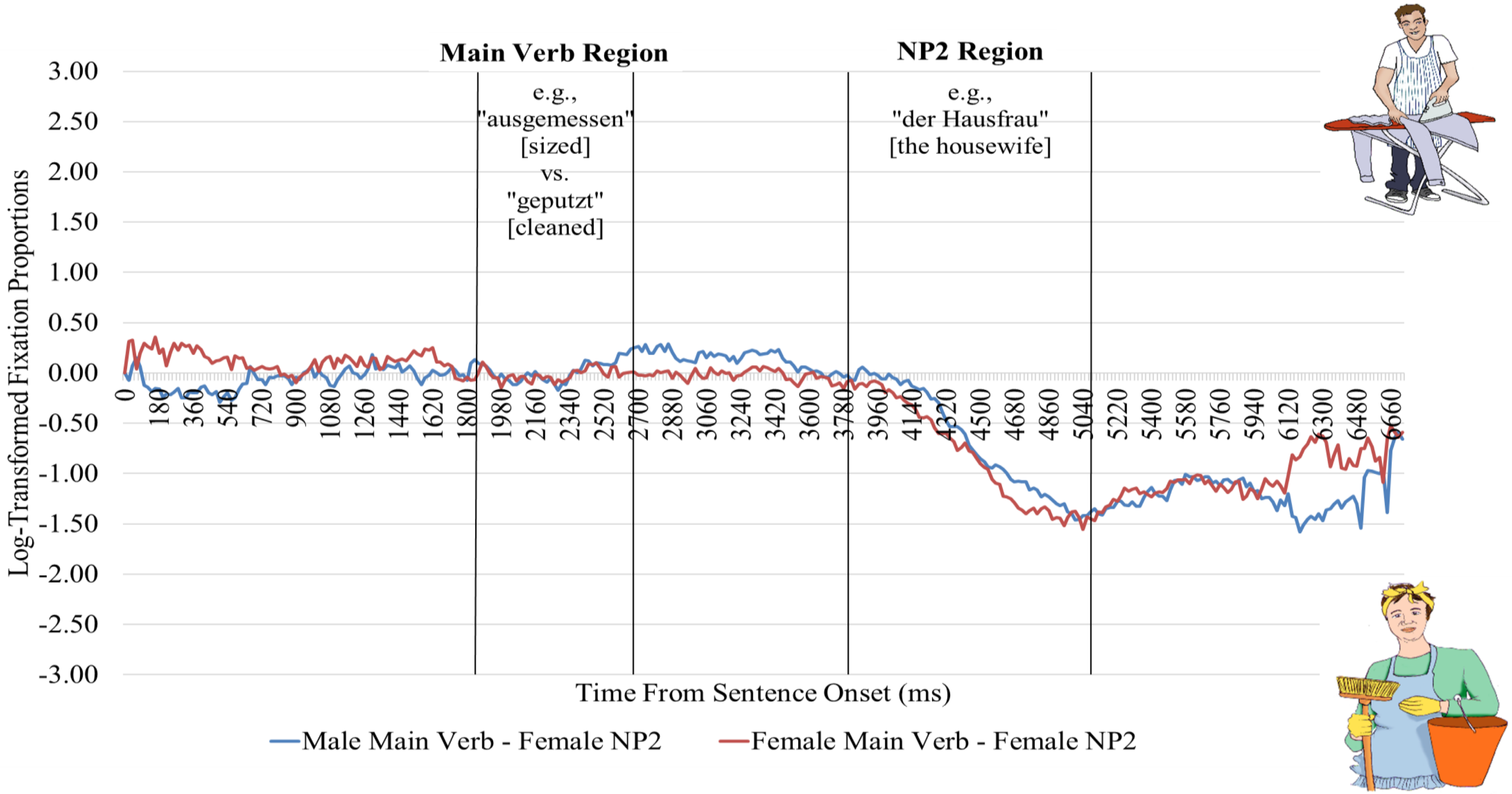


Note. Time course graph depicting changes in log-transformed fixation proportions to the male (positive values) and to the female character (negative values) when encountering sentences with a stereotypically male (blue line) vs. female (red line) main verb referring to a stereotypically male NP2 target. Mean on- and offset times of the main verb region (1. and 2. vertical line) and mean on- and offset times of the NP2 region (3. and 4. vertical line) are indicated (see Table 9.1).

Figure 9.2 illustrates log-ratios for stereotypically male and female main verbs that refer to a female NP2 target: From mean main verb onset up to mean main verb offset, log-ratios ranged mainly about zero. This indicates that, at a descriptive level, both characters were looked at equally likely when listening to the main verbs. Shortly before mean main verb offset until about 200ms before mean NP2 onset, log-ratios for stereotypically male main verbs reached values of about 0.20 which implies that, if at all, there might have been a tendency to look at the male character between about mean main verb offset up to shortly before mean NP2 onset. Regarding the NP2 region, the NP2 target was fixated about 500ms after mean NP2 onset. Compared to time course graphs for sentences ending on a male NP2 target, participants seemingly needed longer to fixate the NP2 target when it was female.

Figure 9.2

Time Course Graph for Sentences With a Stereotypically Male vs. Female Main Verb Referring to a Stereotypically Female NP2 Target



Note. Time course graph depicting changes in log-transformed fixation proportions to the male (positive values) and to the female character (negative values) when encountering sentences with a stereotypically male (blue line) vs. female (red line) main verb referring to a stereotypically female NP2 target. Mean on- and offset times of the main verb region (1. and 2. vertical line) and mean on- and offset times of the NP2 region (3. and 4. vertical line) are indicated (see Table 9.1).

9.2 Experiment 2 – Testing of the Experimental Hypotheses

The main verbs' gender-stereotypicality was hypothesized to lead to fixations on the character whose gender matched the main verbs' gender-stereotypicality (Hypothesis 1). The gender-matching (vs. mismatching) character was hypothesized to be more likely fixated when the main verbs were uttered by a speaker whose gender matched (vs. mismatched) their gender-stereotypicality (Hypothesis 2). Furthermore, participants were expected to more likely fixate the gender-matching character, the higher their values on benevolent and hostile sexism and normative gender-role orientation (Hypothesis 3a). They were expected to less likely fixate the gender-matching character which would imply more looks at the stereotype-inconsistent character, the higher their endorsement of internal and external motivation to control for sexist responses and social desirability (Hypothesis 3b).

To test the hypotheses, data were analyzed and reported analogously to Experiment 1 (see Section 5.2), but with main verb gender-stereotypicality as an experimental factor (see Appendix B, Table B4 for means on the questionnaire measures; Table B5 for Pearson correlations; Table B27 for post-hoc analyses of statistical power for by-participants and by-items data). By-participants data (including questionnaire data) and by-items data were examined for outliers. All participants showed discrete fixation and response patterns. Likewise, none of the items evoked unexpected fixation patterns. Requirements for statistical analyses were proofed to be met before inferential analyses were done.

9.2.1 Experiment 2 – The Effects of Main Verb Gender-Stereotypicality and Speaker Voice on Log-Ratios by Participants (F_1)

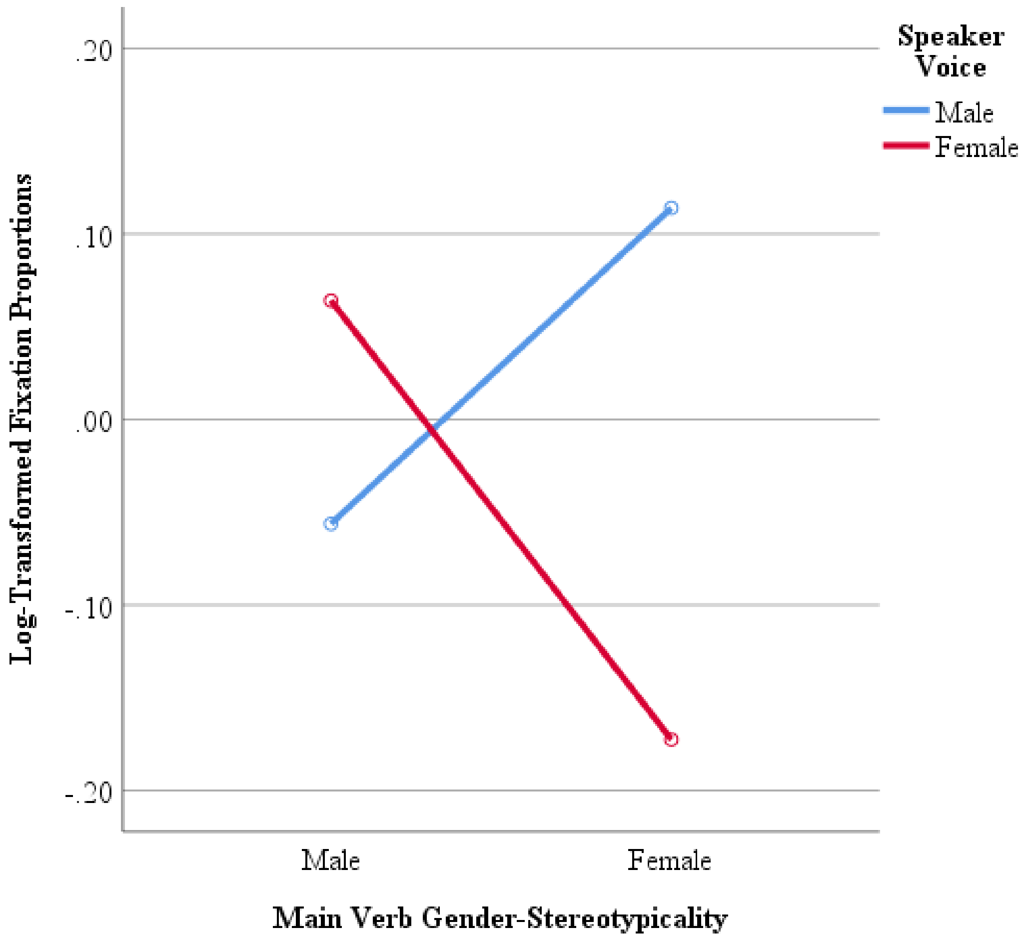
To test Hypothesis 1 and Hypothesis 2 by participants, a repeated measures MANOVA was conducted with log-ratios calculated by participants as a function of main verb gender-stereotypicality as a within-participants factor and speaker voice as a between-participants factor.

The main effects of main verb gender-stereotypicality, sphericity assumed: $F_1(1,76) = 0.04$, $p = .844$, $\eta_p^2 = .001$, and speaker voice, $F_1(1,76) = 0.47$, $p = .495$, $\eta_p^2 = .006$, were not statistically significant in analyses conducted by participants. That is, Hypothesis 1 predicting the main verbs' gender-stereotypicality would guide participants' eye movements to a character whose gender matched the main verbs' gender-stereotypicality was not confirmed in by-participants analyses. The same accounted for Hypothesis 2, predicting a speaker whose gender matched the main verbs' gender-stereotypicality would guide participants' gazes to the gender-matching character. The interaction effect between main verb gender-stereotypicality and speaker voice was not statistically significant, either, sphericity assumed: $F_1(1,76) = 1.47$, $p = .229$, $\eta_p^2 = .019$.

Figure 9.3 visualizes participants' log-ratios as a function of main verb gender-stereotypicality and speaker voice: Although the interaction between main verb gender-stereotypicality and speaker voice was not statistically significant, at a descriptive level, there was a tendency to fixate the character whose gender matched the main verbs' gender-stereotypicality when the main verbs were uttered by a female speaker. Reversely, when the main verbs were uttered by a male speaker, the stereotype-inconsistent character was more likely fixated than the gender-matching one. This fixation pattern was stronger apparent for stereotypically female main verbs than for stereotypically male ones.

Figure 9.3

Log-Transformed Fixation Proportions as a Function of Main Verb Gender-Stereotypicality and Speaker Voice by Participants (F_1)



9.2.2 Experiment 2 – The Effects of Main Verb Gender-Stereotypicality and Speaker Voice on Log-Ratios by Items (F_2)

To test for Hypothesis 1 and Hypothesis 2 by items, a repeated measures MANOVA was performed with log-ratios by items as a function of main verb gender-stereotypicality as a between-items factor and speaker voice as a within-items factor.

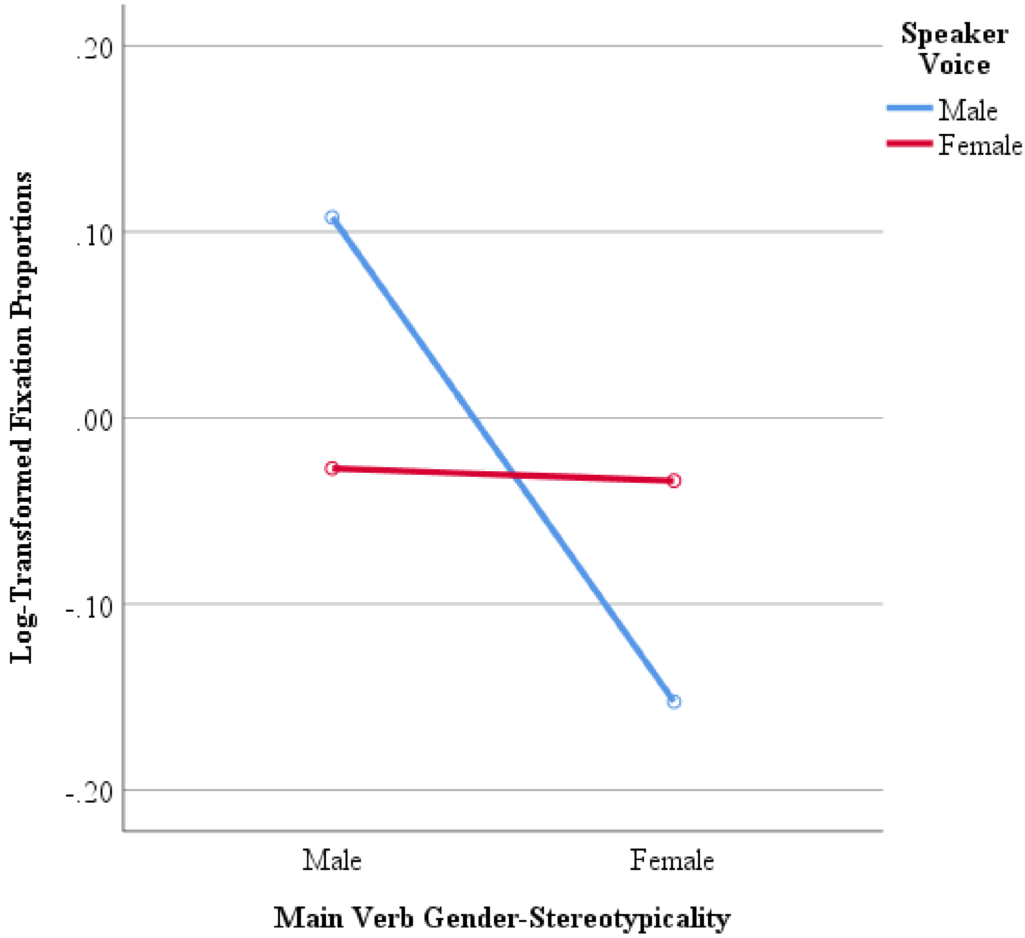
In line with findings by participants, the main effects of main verb gender-stereotypicality, $F_2(1,56) = 1.05$, $p = .311$, $\eta_p^2 = .018$, and speaker voice, sphericity assumed: $F_2(1,56) = 0.01$, $p = .937$, ηp2 < .001, were not statistically significant. That means, Hypothesis 1 predicting the main verbs' gender-stereotypicality would evoke fixations on the character that matches the main verbs' gender-stereotypicality was not confirmed in analyses conducted by items. The same accounted for Hypothesis 2 predicting a speaker whose gender matches (vs. mismatches) the main verbs' gender-stereotypicality would guide participants' gazes to a gender-matching character. The interaction between main verb gender-stereotypicality and speaker voice was not statistically significant, sphericity assumed: $F_2(1,56) = 1.53$, $p = .222$, $\eta_p^2 = .027$.

Figure 9.4 illustrates log-ratios by items as a function of main verb gender-stereotypicality and speaker voice. A slightly different fixation pattern was found by items than by-participants: Items read by a male speaker voice evoked more fixations on the character whose gender matched (vs. mismatched) the main verbs' gender-stereotypicality, particularly when the main verbs were stereotypically female. As indicated by log-ratios slightly below zero, items read by a female speaker voice tended to evoke fixations on the female character, regardless of the main verbs' gender-stereotypicality[30].

[30] Because inferential analyses by participants and by items revealed convergent findings, differences in fixation patterns at a descriptive level should not be overinterpreted. Nevertheless, to rule out errors in the aggregation and restructuring process, by-participants and by-items data were double checked.

Figure 9.4

Log-Transformed Fixation Proportions as a Function of Main Verb Gender-Stereotypicality and Speaker Voice by Items (F_2)



9.2.3 Experiment 2 – The Effects of Main Verb Gender-Stereotypicality, Speaker Voice, and Participants' Gender-Related Attitudes on Log-Ratios

To test Hypothesis 3a and Hypothesis 3b, a repeated measures MANCOVA was performed with log-ratios calculated by participants as a function of main verb gender-stereotypicality as a within-participants factor and speaker voice as a between-participants factor. Participants' endorsement of benevolent and hostile sexism, normative gender role orientation, social desirability, and their internal and external motivation to control for sexist responses were considered as covariates.

In line with analyses of Hypothesis 1 and Hypothesis 2 that did not take into account the covariates, the main effects of main verb gender-stereotypicality, sphericity assumed: $F_1(1,70) = 0.12$, $p = .729$, $\eta_p^2 = .002$, and speaker voice, $F_1(1,70) = 0.80$, $p = .375$, $\eta_p^2 = .011$, were not statistically significant. The same accounted for the interaction between main verb gender-stereotypicality and speaker voice, sphericity assumed: $F_1(1,70) = 2.30$, $p = .134$, $\eta_p^2 = .032$.

Regarding the covariates' impact, the main effects of hostile sexism and normative gender role orientation were statistically significant. The main effects of benevolent sexism, internal and external motivation to control for sexist responses, and social desirability were not statistically significant (see Table 9.2).

Table 9.2

Main Effects of the Covariates on Participants' Log-Ratios Within the Main Verb Region

Covariate	F	p	η_p^2
Constant	0.38	.542	.005
Hostile Sexism	4.28	.042	.058
NGRO	4.28	.042	.058
Internal MCSR	0.14	.708	.002
Benevolent Sexism	0.05	.824	.001
External MCSR	0.03	.872	< .001
Social Desirability	< 0.01	.965	< .001

Note. $df(1,70)$ for all main effects**.** NGRO = Normative Gender Role Orientation, MCSR = Motivation to Control for Sexist Responses[31].

Pearson correlations were calculated to investigate in what direction hostile sexism and normative gender role orientation had affected participants' eye gazes. They revealed that the higher participants' levels of hostile sexism, the more they looked at the female character when listening to stereotypically male main verbs, $r(76) = -.15$, $p = .206$, and stereotypically female main verbs, $r(76) = -.05$, $p = .665$. The higher participants' levels of normative gender role orientation, the more they looked at the female character when listening to stereotypically male main verbs, $r(76) = -.01$, $p = .927$, and at the male character when listening to stereotypically female main verbs, $r(76) = .12$, $p = 314$. The interaction effects between main verb gender-stereotypicality and any of the covariates were not statistically significant (see Table 9.3).

[31] The main effects of hostile sexism and normative gender role orientation were exactly the same. Therefore, the data were double checked to make sure everything is correct.

Table 9.3

Interaction Effects Between Main Verb Gender-Stereotypicality and the Covariates

Main Verb Gender-Stereotypicality x Covariate	F	p	η_p^2
… External MCSR	2.57	.113	.035
… Benevolent Sexism	0.96	.331	.013
… NGRO	0.69	.409	.010
… Hostile Sexism	0.14	.707	.002
… Internal MCSR	0.07	.793	.001
… Social Desirability	0.06	.805	.001

Note. $df(1,70)$ for all interaction effects. NGRO = Normative Gender Role Orientation, MCSR = Motivation to Control for Sexist Responses.

Following the principle of parsimony (see Tabachick & Fidell, 2007), a similar MANCOVA was performed with only hostile sexism and normative gender role orientation as covariates. The results were in line with the analysis in which all covariates were included. The statistically significant main effects of hostile sexism, $F_1(1,74) = 6.21$, $p = .015$, $\eta_p^2 = .077$, and normative gender role orientation, $F_1(1,74) = 5.42$, $p = .023$, $\eta_p^2 = .068$, were confirmed.

That is, participants' endorsement of hostile sexism resulted in more looks at the female character independent of the main verbs' gender-stereotypicality. Participants' endorsement of normative gender role orientation led to more looks at the stereotype-inconsistent character. Hypothesis 3a predicting participants would more likely fixate the character whose gender matched (vs. mismatched) the main verbs' gender-stereotypicality, the higher their levels of benevolent and hostile sexism and normative gender role orientation was thus not confirmed. The same accounted for Hypothesis 3b predicting participants would less likely fixate the gender-matching (vs. mismatching) character, the higher their levels of external and internal motivation to control for sexist responses and social desirability (see also Appendix B, Table B4 for participants' mean scores; Table B5 for the full set of Pearson correlations, Table B27 for post-hoc analyses of statistical power by participants and by items).

9.3 Experiment 2 – Manipulation Check

Inferential analyses of the NP2 region served as a manipulation check to confirm that the NP2 target was looked at independent of whether it was specified by a male or a female speaker voice. Inferential analyses of the NP2 region were done analogously to Experiment 1.

9.3.1 Experiment 2 – The Effects of NP2 Gender and Speaker Voice on Log-Ratios by Participants (F_1)

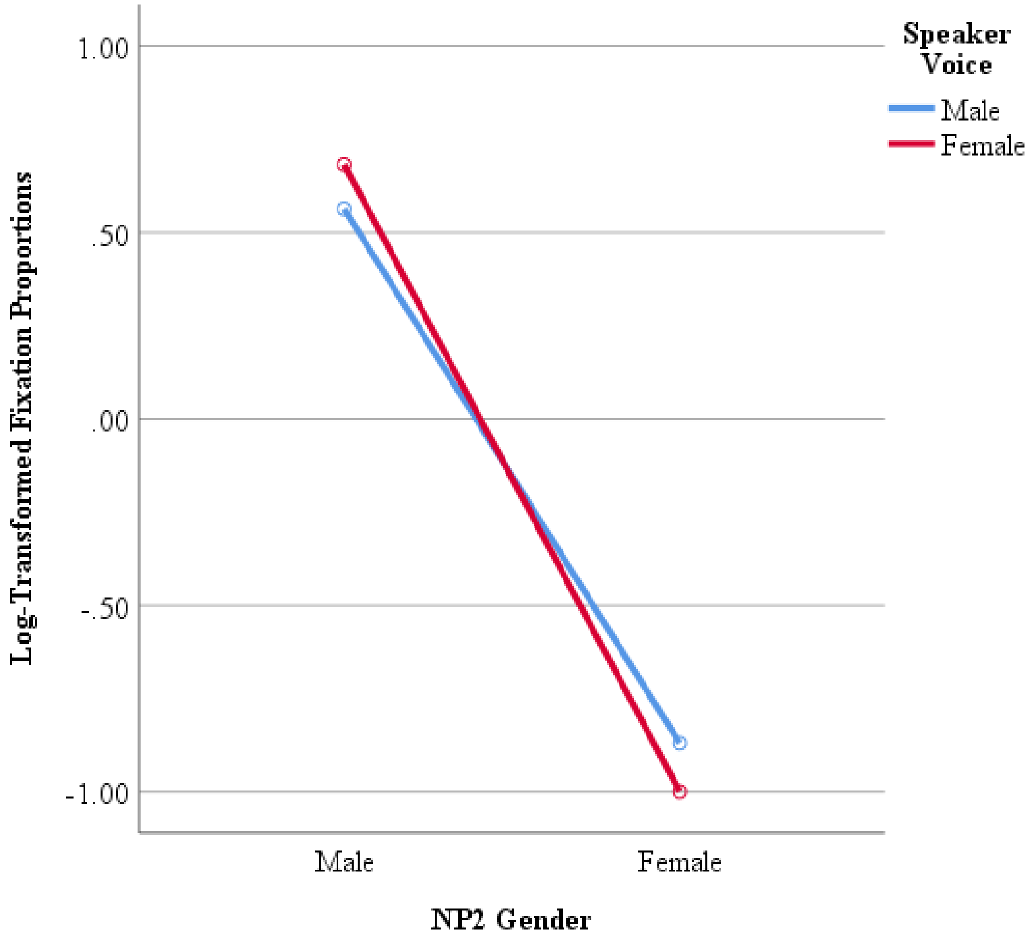
A repeated measures MANOVA was performed with log-ratios calculated by participants as a function of NP2 gender as a within-participants factor and speaker voice as a between-participants factor.

The main effect of NP2 gender was statistically significant in analyses conducted by participants, sphericity assumed: $F_1(1,76) = 148.75$, $p < .001$, $\eta_p^2 = .662$. That is, in line with the time course graphs, the NP2 target was looked at when it was mentioned at the end of the sentence. The main effect of speaker voice, $F_1(1,76) < 0.01$, $p = .949$, $\eta_p^2 < .001$, and the interaction between NP2 gender and speaker voice, sphericity assumed: $F_1(1,76) = 0.95$, $p = .332$, $\eta_p^2 = .012$, were not statistically significant.

Figure 9.5 illustrates log-ratios by participants as a function of NP2 gender and speaker voice: As indicated by the statistically significant main effect of NP2 gender and the statistically non-significant main effect of speaker voice, the NP2 target was looked at independent of whether it was uttered by a male or a female speaker. However, at a descriptive level, the NP2 target tended to be more likely looked at when it was uttered by a female (vs. male) speaker. Moreover, log-ratios for male NP2 targets reached values of about .50, while log-ratios for female NP2 targets ranged around -1.00 which indicates that, independent of the speaker voice, the NP2 target was more likely fixated when it was female (vs. male).

Figure 9.5

Log-Transformed Fixation Proportions as a Function of NP2 Gender and Speaker Voice by Participants (F_1)



9.3.2 Experiment 2 – The Effects of NP2 Gender and Speaker Voice on Log-Ratios by Items (F_2)

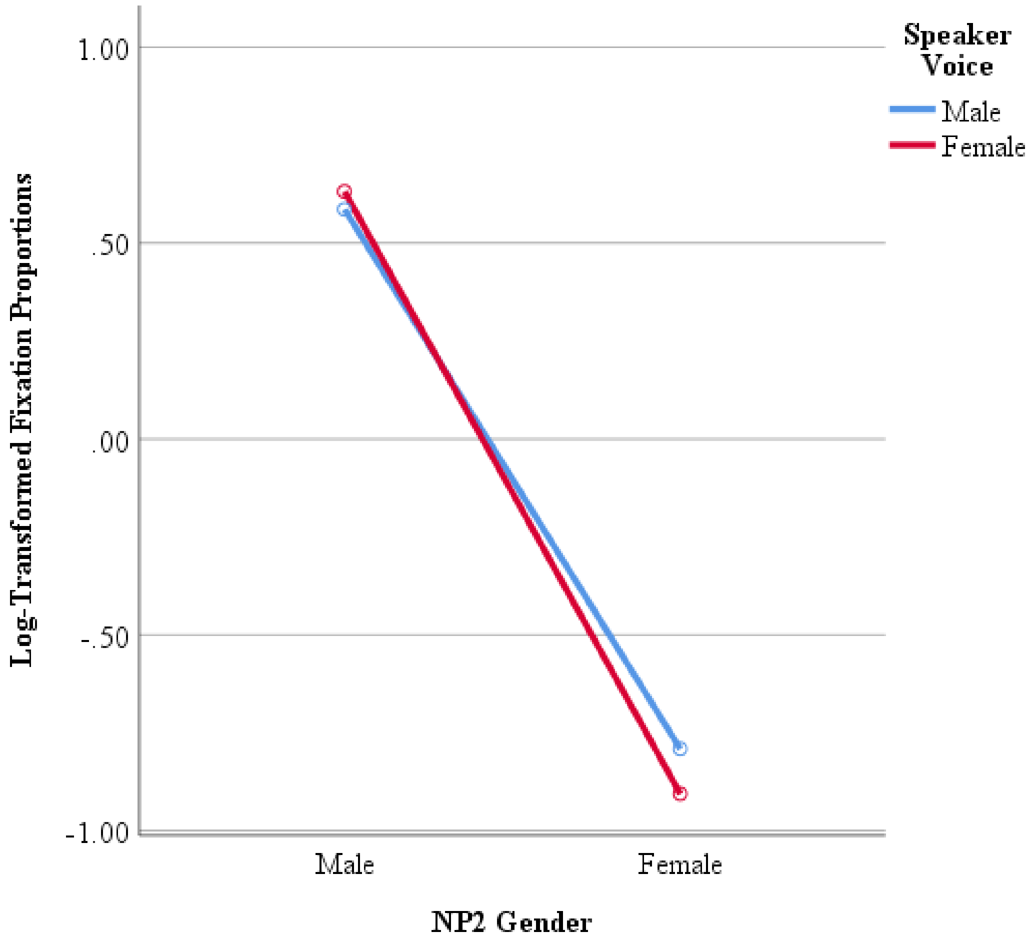
To test the effects of NP2 gender and speaker voice by items, a repeated measures MANOVA was computed with log-ratios calculated by items as a function of NP2 gender as a between-items factor and speaker voice as a within-items factor.

Confirming by-participants analyses, the main effect of NP2 gender was statistically significant, $F_2(1,56) = 327.29$, $p < .001$, $\eta_p^2 = .854$, in analyses conducted by items. That is, the NP2 target was fixated when it was named. The main effect of speaker voice, sphericity assumed: $F_2(1,56) = 0.17$, $p = .681$, $\eta_p^2 = .003$, was not statistically significant. This indicates that the NP2 target was looked at, irrespective of whether it was specified by a male or a female speaker. The interaction between NP2 gender and speaker voice, sphericity assumed: $F_2(1,56) = 0.91$, $p = .345$, $\eta_p^2 = .016$, was not statistically significant, either.

Figure 9.6 visualizes log-ratios by items as a function of NP2 gender and speaker voice: In line with by-participants analyses, the NP2 target was looked at, irrespective of the speaker voice. At a descriptive level, the tendency to more likely fixate the NP2 target when listening to a female speaker that was found for fixation patterns by participants, was confirmed by items. Likewise, log-ratios' range for male NP2 targets and for female NP2 targets indicated that the NP2 target was more likely looked at when it was female (vs. male).

Figure 9.6

Log-Transformed Fixation Proportions as a Function of NP2 Gender and Speaker Voice by Items (F_2)



9.4 Experiment 2 – Exploratory Analyses

Similar to Experiment 1, exploratory analyses were conducted to explain the results in more detail (see Appendix B, pp. 315).

10 Experiment 2 (Main Verbs - Humans) – Discussion

Experiment 2 was analogous to Experiment 1. The experimental setup and data analyses were thus similar to Experiment 1, except that the main verbs' gender-stereotypicality was in focus of investigation.

Contrary to Hypothesis 1, the main effect of main verb gender-stereotypicality was not statistically significant. That is, the main verbs' gender-stereotypicality did not guide participants' eye movements to a character whose gender matched the main verbs' gender-stereotypicality. Contrary to Hypothesis 2, the main effect of speaker voice was not statistically significant, either. That is, participants' eye movements were not guided to the gender-matching character when the speaker's gender matched (vs. mismatched) the main verbs' gender-stereotypicality. Similarly, the interaction effect between main verb gender-stereotypicality and speaker voice was not statistically significant. This was found consistently for data analyzed by participants and by items. The finding that neither the main effect of main verb gender-stereotypicality nor the interaction between main verb gender-stereotypicality and speaker voice guided participants' eye movements in Experiment 2 suggests that the main verbs may not be immediately linked to male or female gender. This may have been due to the fact that gender-stereotypical tasks and roles have changed drastically over the past decades (see Auster & Ohm, 2000; Eagly et al., 2019). That way, the lines between stereotypically male and female tasks and roles got blurred. Because people infer an individual's attributes from her or his tasks and roles (Eagly et al., 2019; Koenig & Eagly, 2014; Wood & Eagly, 2012), changed gender roles and tasks need to precede changed trait attributions to men and women. In turn, associations between attributes and gender may persist for a longer period of time and thus be stronger than associations between roles or tasks and gender. This could explain why, in Experiment 1, the adverbs apparently automatically activated associations with male and female gender which may have facilitated to link them to either of the characters. While in Experiment 2, the main verbs may have been less strongly associated with gender. Therefore, it seems that in Experiment 2, participants first needed to comprehend that either of the characters was referred to. This likely needed more time and cognitive capacities than in Experiment 1 (see e.g., Fazio, 2007; Schneider & Chein, 2003 for the role of cognitive capacities). Because some aspects remained constant across trials, such as the speaker voice, the sentence's structure, and the array of the visual scenes, participants may have sensed that the sentences may be related to gender stereotypes after a while. Similar to Experiment 1, participants had indicated low levels of sexism and relatively high levels of internal motivation to control for sexist responses and social desirability. Therefore, the idea to encounter gender stereotypes may have caused fear and discomfort to possibly behave in a gender-stereotypical or sexist manner (see Monteith, 1993; Monteith et al., 2002). However, given the assumption that it was difficult to infer the sentence's target from the main verbs, feelings of discomfort likely arose later and less intense than in Experiment 1. That way, participants' vigilance for aspects that could be related to gender stereotypes and sexism may have been less strong than in Experiment 1. As a consequence, participants in Experiment 2 might have attempted to counter gender stereotypes and sexism later and less efficiently than participants in Experiment 1.

This might explain the results obtained when Hypothesis 3a and Hypothesis 3b were tested. More specifically, the main effects of participants' endorsement of hostile sexism and normative gender role orientation statistically significantly affected their visual attention when listening to the main verbs. Pearson correlations were performed to explore in what sense the covariates had affected participants' fixations: Contrary to Hypothesis 3a, the more hostile sexist attitudes participants reported, the more they looked at the female character independent of the main verbs' gender-stereotypicality. The higher participants' endorsement of normative gender role orientation, the more they looked at the stereotype-inconsistent character. Contrary to Hypothesis 3b, participants' internal and external motivation to control for sexist responses and social desirability did not guide participants' visual attention when listening to the main verbs. Given the assumption that participants have noticed that the sentences always contained an action performed by a male or a female character across trials, the main effects of hostile sexism and normative gender role orientation might have been due to the constructs' focus on men's and women's roles. To illustrate, items on hostile sexism contain abstract verb constructs and metaphors to describe women's actions (e.g., "Once a woman <u>gets a man to commit to her</u>, she usually tries to <u>put him on a tight leash</u>."). Therefore, the higher participants endorsement of hostile sexism, the more they may have expected the female character to be the sentence's target. Assuming that participants had difficulty to predict the sentence's target from the main verbs' gender-stereotypicality, this expectation may have helped them to comprehend the sentences. Items on normative gender role orientation contain concrete main verbs about what men and women stereotypically do (e.g., "<u>To iron shirts</u> is not men's business."). This is similar to the main verbs used in Experiment 2. Therefore, the more participants, endorsement on normative gender role orientation, the more they might have been sensitive for the main verbs' gender-stereotypicality. This might have facilitated to link the main verbs to the characters and finally to avoid fixations on the character whose gender matched the main verbs' gender-stereotypicality which resulted in more fixations on the stereotype-inconsistent character.

Moreover, explorations of gender effects revealed a statistically significant triple interaction between main verb gender-stereotypicality, speaker voice, and participant gender: Participants more likely looked at the character whose gender matched (vs. mismatched) the main verbs' gender-stereotypicality when the main verbs were articulated by a speaker of their own gender. They more likely looked at the stereotype-inconsistent (vs. gender-matching) character when the main verbs were uttered by a speaker of their opposite gender. This suggests that participants' and the speaker's gender may have played a role when listening to gender-stereotypical main verbs. This triple interaction between main verb gender-stereotypicality, speaker voice, and participant gender, however, turned out just not statistically significant in an analysis in which furthermore the effects of hostile sexism and normative gender role orientation were considered that were found to have affected participants' gazes in the context of Hypothesis 3a. In this exploratory analysis, also the main effect of normative gender role orientation turned out just not statistically significant. Only the statistically significant main effect of hostile sexism was confirmed. This suggests that participants' endorsement of hostile sexism, rather than normative gender role orientation and the triple interaction between main verb gender-stereotypicality, speaker voice, and participant gender may have determined which character was looked at when listening to the main verbs. This appears plausible given the assumption that hostile sexism's focus on women's actions was used to predict the sentence's target, while counteractive fixations on the stereotype-inconsistent character were possible only after the main verbs could be linked to either of the characters. These findings may thus strengthen the assumption that participants had difficulty to draw a link between the main verbs and the characters which thus required time and cognitive capacities. However, as soon as participants had been aware of gender-stereotypes and sexism, they apparently attempted to counter the main verbs' gender-stereotypicality. Similar to Experiment 1, participants' and the speaker's group membership in terms of gender may have played a role while doing so.

In Experiment 2, participants apparently needed the entire sentence to comprehend which of the characters was the sentence's target. To illustrate, a statistically significant main effect of the NP2 gender indicated that participants looked at the sentence's target when its gender was explicitly specified at the end of the sentence. The main effect of speaker voice and the interaction between NP2 gender and speaker voice were not statistically significant. That is, participants looked at the NP2 target independent of the speaker's gender. However, time course graphs demonstrated that the NP2 target was fixated shortly after its mean onset time. In Experiment 1, the NP2 target was fixated at its mean onset and even slightly before when it was female. Participants' fixation pattern at the end of the sentence thus suggests, the more difficult it was to link the main verbs to the characters, the more difficult it seemed to identify the sentence's target when it was named. Regarded in the context of the Social CIA (Münster & Knoeferle, 2018) this appears plausible because an initial link between the main verbs and the characters serves to predict further incoming information. Moreover, descriptive analyses showed that participants were more likely to look at the NP2 target when it was female (vs. male). This tendency was also found in Experiment 1 and in Guerra, et al. (2021). It may be due to the fact that in German language, the suffix "-in" clearly marks a female target, while the male form can be generically used for male and female targets.

In sum, it seems that, unlike the adverbs in Experiment 1, the main verbs could not be immediately linked to male or female gender. That way, participants apparently first needed to comprehend to which character the main verbs referred. This however seemed to require time and cognitive capacities. However, participants seemingly attempted to counter gender stereotypes and sexism as soon as they had comprehended that the sentences referred to male or female targets and that the sentences' content might thus be related to gender stereotypes. Participants' and the speaker's group membership in terms of gender may have played a role while doing so.

11 Experiment 3 (Adverbs – Robots) – Hypotheses

Experiment 3 was an extended replication of Experiment 1. The main difference between the two experiments was that human targets were replaced by robot targets.

1 the adverbs' gender-stereotypicality and

2 a speaker voice whose gender matches (vs. mismatches) the adverbs' gender-stereotypicality

lead to higher log-gaze probabilities to look at a robot whose gender matches (vs. mismatches) the adverbs' gender-stereotypicality[1]

Log-gaze probabilities to look at a robot whose gender matches (vs. mismatches) the adverbs' gender-stereotypicality are <u>higher</u>, the higher participants' values on

3a Benevolent and hostile sexism and normative gender role orientation

and <u>lower</u>, the higher participants' values on

3b Internal and external motivation to control for sexist responses and social desirability.

For exploratory purposes, furthermore participants' perceptions of and attitudes toward robots (i.e., robot human- and machinelikeness, robot agency, robot communion, robot acceptance, robot anxiety), their tendency to anthropomorphize non-human agents, technology commitment, and their evaluations of the presented verbal and visual stimuli (i.e., the robot stimuli's recognizability in terms of gender and profession, the robot professions' typicality, imaginability of robots performing the presented tasks and professions, and robot use preferences) were investigated. This allowed to gain insights into participants' perceptions of robots in general and of the presented robot stimuli in particular and to furthermore explore whether participants' perceptions of robots affected their visual attention.

12 Experiment 3 (Adverbs – Robots) – Method

12.1 Pretest V – Identification of 'Typical' Robots

To investigate participants' visual attention when robots were displayed, the human characters that represented the gender-stereotypical professions in Experiment 1 and Experiment 2 had to be replaced by robots.

The major aim of Pretest V was thus to identify a set of robots that could be displayed as male and female representatives of the gender-stereotypical professions. To be suitable for Experiment 3, 'real' existing robots were needed that were used in research contexts or for commercial purposes. This was important to mirror participants' associations between spoken language and real robots, instead of any fantasy or 'science fiction-like' image of them. Furthermore, the robots should be judged as 'typical' exemplars of the category 'robot'. People commonly do not yet have much experience with robots (see Bernotat & Eyssel, 2018; Bernotat et al.,2021). Typicality was thus a major requirement of the robot drawings because it likely enhanced the recognizability of the sentence's target as being actually a robot instead of any drawing of a humanlike figure.

It was deemed challenging to portray the robots as male and female representatives of a gender-stereotypical profession. Therefore, Pretest V furthermore served to explore which robot could be easily adapted in terms of gender, what could make it appear as male or female, and which profession each of the robots could best represent.

12.1.1 Pretest V – Material: Robot Platforms

First, a set of 14 robots was preselected that could be deemed typical robots and that were either used in research contexts or that were available for commercial use. To have optimal and standardized images, colored portraits of the robots were drawn[32] (see Appendix A, pp. 277; Bernotat et al.,2017; 2021 for similar stimuli).

12.1.2 Pretest V – Procedure

Pretest V was done in paper-pencil form. Being told that I would be interested in their personal impression of a new robot design, participants judged either of the 14 robots. Participants were instructed to carefully watch the robot that was pictured on top of each page

[32] Janik Sachse, a former B.Sc.-student, created all the robot drawings that were used in this set of experiments. For some robots, an existing robot head had to be combined with a body of another existing robot because some robots do not have a body (e.g., *FloBi* robot).

of the questionnaire (10.27cm x 7.78cm) before answering the questions related to the robot's design (see Appendix A, p. 276 for Pretest V instructions).

Using 7-point Likert scales, participants were asked to indicate to what extent they perceived the robot as being *recognizable* and as a *typical* exemplar of the category robot. As in Pretest III (see Section 4.3), Likert scales ranged from 1 (= not at all) to 7 (= very much) for all items. To get more detailed information, participants were provided open response formats to explain which characteristics of the depicted robot they perceived as especially *typical* or as *untypical*, respectively and what changes they would suggest to make the robot seem more typical. Then, participants judged to what degree they perceived the robot as male and as female using 7-point Likert scales. Open response formats were given to justify their judgements.

To find out which profession each robot could best represent, the set of professions was listed (see Appendix A, Table A17). Participants used 7-point Likert scales to indicate to what extent they deemed the robot suitable for each of the listed professions. To avoid any bias because participants knew the robot from other contexts and to make sure they believed Pretest V was about the evaluation of a new robot design, participants were asked whether they were familiar with the depicted robot. If so, they had to specify from what context they knew the robot and whether they had already participated in another study using this or any other robot. Finally, they reported demographics (i.e., age, professional status, and gender). Completing the questionnaire took about ten minutes. Participants got 0.5 course credits when needed.

12.1.3 Pretest V – Sample

In total, $N = 394$ participants[4] completed the questionnaire. 55 respondents had to be excluded from data analyses because they indicated to have been familiar with the displayed robot before (e.g., from media, other studies, or their work context). Thus, their ratings of the robot's recognizability and typicality might have been biased. Most of the remaining $n = 339$ (male: $n = 112$, female: $n = 222$, 'genderfluid': $n = 2$, non-binary: $n = 1$, undisclosed: $n = 2$; $M_{age} = 22.55$, $SD_{age} = 5.25$, age range: 15-73 years) were students (student: $n = 317$, professional: $n = 6$, undisclosed: $n = 16$).

12.1.4 Pretest V – Results

First, following the major aim of Pretest V to find out which of the robots were suitable to be used in Experiment 3, one-sample t-tests against the scale midpoint of 4 were performed to test whether the robots were recognized as a robot and whether they were perceived as a typical robot. All robots were recognized as such. Except for the *Floka 2* robot, all robots were furthermore judged as typical robots (see Table 12.1). Therefore, except for *Floka*

2 robot, all robots were suitable to be used for further adaptation in terms of gender and profession (see also Appendix A, pp. 281 for participants' indications of what characteristics they considered typical or untypical for a robot).

Table 12.1

Results of One-Sample t-Tests Against the Scale Midpoint of 4 on Participants' Evaluations of the Robots as Being Recognizable and as Typical

Robot	Item	*M*	*SD*	*df*	*t*	*p*	Cohen's *d*
1. ASIMO	Recognizable	6.07	1.54	26	6.99	< .001	1.34
	Typical	5.52	1.40	26	5.65	< .001	1.09
2. Floka	Recognizable	6.64	0.49	24	26.94	< .001	5.39
	Typical	4.96	1.17	24	4.10	< .001	0.82
3. Floka 2	Recognizable	6.74	0.62	22	21.22	< .001	4.42
	Typical	4.70	1.72	22	1.94	.065	0.41
4. Folkwang	Recognizable	6.63	0.49	23	26.00	< .001	5.31
	Typical	5.92	0.93	23	10.11	< .001	2.06
5. iCub	Recognizable	6.95	0.22	20	62.00	< .001	13.53
	Typical	5.76	0.89	20	9.08	< .001	1.98
6. Meka	Recognizable	6.13	1.30	23	8.03	< .001	1.64
	Typical	5.29	1.55	23	4.09	< .001	0.84
7. Myon	Recognizable	6.92	0.28	24	52.73	< .001	10.55
	Typical	5.84	0.85	24	10.82	< .001	2.16
8. NAO	Recognizable	6.32	0.77	27	15.90	< .001	3.01
	Typical	5.36	1.39	27	5.15	< .001	0.97
9. Pepper	Recognizable	6.86	0.52	28	29.88	< .001	5.55
	Typical	5.55	1.12	28	7.46	< .001	1.38
10. Ri-Man	Recognizable	6.67	0.56	23	23.14	< .001	4.72
	Typical	5.33	1.34	23	4.87	< .001	0.99
11. RoboThespian	Recognizable	6.80	0.41	19	30.51	< .001	6.82
	Typical	5.75	0.64	19	12.25	< .001	2.74
12. Romeo	Recognizable	6.85	0.60	26	24.64	< .001	4.74
	Typical	5.96	1.13	26	9.06	< .001	1.74
13. SociFlobot	Recognizable	6.76	0.44	20	29.00	< .001	6.33
	Typical	5.38	1.36	20	4.66	< .001	1.02
14. SociFlobot 2	Recognizable	6.25	1.33	19	7.55	< .001	1.69
	Typical	4.70	1.17	19	2.67	.015	0.60

To explore participants' judgements of the robot's gender, independent *t*-tests relative to the scale midpoint of 4 were performed on participants' mean ratings of a respective robot as male and female. Apart from only a few exceptions, the robots were generally rather judged as male than as female (see Table 12.2).

Table 12.2

Results of One-Sample t-Tests Against the Scale Midpoint of 4 on Participants' Evaluations of the Robots as Male and as Female

Robot	Item	*M*	*SD*	*df*	*t*	*p*	Cohen's *d*
1. ASIMO	Male	5.74	1.43	26	6.32	< .001	1.22
	Female	2.15	1.32	26	-7.28	< .001	1.40
2. Floka	Male	4.48	1.36	24	1.77	.090	0.35
	Female	2.84	1.37	24	-4.22	< .001	0.84
3. Floka 2	Male	4.17	1.53	22	0.55	.590	0.11
	Female	2.68	1.62	21	-3.83	.001	0.82
4. Folkwang	Male	5.25	1.73	23	3.55	.002	0.72
	Female	2.25	1.26	23	-6.81	< .001	1.39
5. iCub	Male	4.10	1.59	19	0.28	.781	0.06
	Female	3.35	1.60	19	-1.82	.085	0.41
6. Meka	Male	4.96	1.63	23	2.88	.008	0.59
	Female	1.95	1.17	21	-8.17	< .001	1.74
7. Myon	Male	4.68	1.60	24	2.13	.044	0.43
	Female	2.64	1.47	24	-4.63	< .001	0.93
8. NAO	Male	4.79	1.47	27	2.82	.009	0.53
	Female	2.96	1.53	27	-3.59	.001	0.68
9. Pepper	Male	4.28	1.81	28	0.82	.419	0.15
	Female	2.86	1.64	28	-3.73	.001	0.69
10. Ri-Man	Male	5.24	1.61	24	3.84	.001	0.77
	Female	1.75	1.07	23	-10.27	< .001	2.10
11. RoboThespian	Male	4.80	1.61	19	2.22	.039	0.50
	Female	2.25	1.16	19	-6.72	< .001	1.50
12. Romeo	Male	5.00	1.52	26	3.42	.002	0.66
	Female	2.65	1.35	25	-5.07	< .001	0.99
13. SociFlobot	Male	4.62	1.56	20	1.81	.085	0.40
	Female	3.43	1.78	20	-1.47	.156	0.32
14. SociFlobot 2	Male	4.60	1.60	19	1.67	.110	0.37
	Female	2.35	1.53	19	-4.82	< .001	1.08

Likewise, participants' mean ratings of each of the 14 depicted robot's suitability for the gender-stereotypical professions were calculated relative to the scale midpoint of 4. The table in which the results were reported is relatively long. Therefore, for the sake of readability, the results are listed in Appendix A, Table A31. Participants' ratings served as 'guidelines' to assign robots to a specific gender-stereotypical profession. They were thought to possibly facilitate the later adaptation process of the robot drawings in terms of gender and profession.

Nonetheless, because the adaptation of the robot drawings needed to be done by an external designer whose time was limited, the time it might take to adapt the respective robots needed to be considered, too. To illustrate, due to the number of single cables and technical details, to adapt, which required to newly draw, the robot *RoboThespian* was expected to be more difficult and to take longer than to adapt *Pepper* robot. Therefore, *RoboThespian* was one of the robots that were not selected for further adaptation although it was recognized as a typical robot.

Seven robots were finally selected for further adaptation. To present each robot equally often in Experiment 3 and to balance its profession's gender-stereotypicality, each robot was assigned to two gender-stereotypical professions, a stereotypically male and a stereotypically female one (see Table 12.3).

Table 12.3

Robots That Were Selected for Further Adaptation in Terms of Robot Gender and Gender-Stereotypical Profession and the Professions They Were Assigned to

Robot	Profession	Profession's Gender-Stereotypicality
1. ASIMO	Bouncer	Male
	Perfumery Shop Assistant	Female
2. Folkwang	Soldier	Male
	Cosmetician	Female
3. Myon	Gold Digger	Male
	Florist	Female
4. NAO	Firefighter	Male
	Stewardess	Female
5. Ri-Man	Construction Worker	Male
	Dental Assistant	Female
6. Romeo	Mechanics Engineer	Male
	Housewife	Female
7. SociFlobot	Butcher	Male
	Hair Dresser	Female

12.2 Pretest VI – Evaluation of the Robot Drawings in Terms of Gender and Profession

The robots that were selected based on Pretest V (see Section 12.1) were adapted in terms of gender and profession. Because the recognizability of the robot stimuli was crucial to draw a link between spoken language and the visual scenes in Experiment 3, Pretest VI served to evaluate whether the final robot drawings were recognized as male and female representatives of their assigned profession. To enhance the recognizability of the robots' professions, the robots should also be regarded as representing their assigned profession in a typical manner.

12.2.1 Pretest VI – Robot Drawings

A set of 14 pairs of male and female robots that represented one of the gender-stereotypical professions was created (see Appendix A, pp. 286). The two robots of a pair were identical except for their robot gender. Keeping the robots' design constant within robot pairs was crucial to avoid that one robot of a pair would attract more visual attention due to its design than the other. Following Bernotat et al. (2017; 2021), robot gender was induced by manipulating the robots' waist-to-hip-ratio (WHR) and shoulder width (SW): 'Male' robots were designed with a WHR of 0.9 and 100% SW relative to the 'female' robots. Reversely, female robots had a WHR of 0.5 and 80% SW relative to their male counterparts. In case the robots had a hair part (i.e., *Folkwang* and *SociFlobot* robots), the male robot was pictured with a short hair part, while the female robot was shown with a long hair part (see Eyssel & Hegel, 2012). To present the gender-stereotypical professions, the respective profession was symbolized on the robot's torso**.** Furthermore, the robots were colored in a manner that was commonly known as being typical for their assigned profession when possible.

12.2.2 Pretest VI – Procedure

Pretest VI was conducted using *Qualtrics* (*Qualtrics*^{XM}), a tool for online-surveys. Participants were informed that they would be presented with a set of robot pairs that each depicted a certain profession. Respondents' task was to guess which profession each robot pair might represent (see Appendix A, p. 285 for Pretest VI instructions). I opted to display the male and the female robot as a pair because the robots were to be presented together in Experiment 3. Only one robot pair was presented at a time. It was displayed on top of each page of the questionnaire.

First, after having looked at the robot pair carefully, participants had to guess which profession the robot pair might represent. This served to assess whether participants identified the robots' profession at a first view when no further information about the robots' profession was given. An open response format was provided not to bias participants' responses. In the following, the same robot pair was presented again. Unlike before, participants were asked to solely evaluate either the male or the female robot. The robot that was to be evaluated was circled in red. The male robot of a pair was judged first. Using a 7-point bipolar scale, participants indicated whether they perceived the robot as *male* (=1) or as *female* (= 7). The scale midpoint of 4 marked perceived gender-neutrality. Then, similar to Pretest V, using 7-point Likert scales, respondents indicated to what degree they deemed the robot as *recognizable* and as representing the assigned profession in a *typical* manner (1 = not at all, 7 = very much). It was pointed out that the typicality of the profession's depiction was to be evaluated, not participants' personal beliefs whether they considered robots as typical representatives of the portrayed profession in general. The depicted profession was explicitly named in the items. This was done because in Experiment 3, the robots were presented on the screen while the assigned profession was named. This way, I could investigate whether participants linked the displayed robot to the profession when it was mentioned. After having judged the male robot, the female robot of the same robot pair was to be judged analogously. Then, the next robot pair was presented and so forth.

After having evaluated all robot pairs, some questions followed to explore participants' beliefs about robots performing professions in general and to assess their prior experience with robots: Participants indicated to what extent they could imagine robots in general to perform the professions that were named in Pretest VI (1 = not at all, 7 = very much). Using open response formats, it was enquired which professions participants could imagine robots to perform and which not and whether they had known one of the pictured robots before. Finally, participants indicated demographics (i.e., age, professional status, and gender) and received 0.5 course credits if needed. Completing the questionnaire took about 15 minutes.

12.2.3 Pretest VI – Sample

$N = 36$ participants[4] completed Pretest VI (male: $n = 9$, female: $n = 20$, non-binary: $n = 1$, undisclosed: $n = 6$; $M_{age} = 25.47$, $SD_{age} = 13.31$, age range: 14-81 years). Most respondents indicated to be students of German nationality (student: $n = 19$, undisclosed: $n = 17$; German nationality: $n = 29$, other nationality: $n = 1$, undisclosed: $n = 6$). All of them had a good command of German language (German native speaker: $n = 28$, other native language: $n = 2$, undisclosed: $n = 6$). Seven participants indicated having been familiar with one of the robots before[33]: Four of them told they had known one of the robots from movies, books, and video

[33] Unlike in Pretest V (Section 12.1) participants who had indicated to have been familiar with one of the robots before had not to be excluded from data analyses because prior knowledge of the robots was unlikely to bias participants' evaluations of the robots in terms of gender, the recognizability of the displayed professions, and the professions as being portrayed in a typical manner.

games. This however cannot be true because I only selected robots that were used in research or for commercial purposes, but no robots from science-fiction stories (see Section 12.1). Three participants indicated to have known Pepper and NAO robots from studies and expositions.

12.2.4 Pretest VI – Results

Regarding participants' spontaneous guesses about the professions portrayed by the robot pairs, the majority of the participants correctly identified most of the robots' professions from a first view. The number of the correct identifications was counted (see Table 12.4). The robot pairs that portrayed a male and a female dental assistant, mechatronics engineer, and gold digger received the least correct identifications. This indicates that participants had difficulties to recognize the robots' assigned profession when the profession was not explicitly mentioned. Regarding the depiction of the dental assistants, all participants considered the robot pair as portraying a male and a female dentist. Therefore, the number of correct answers is zero for this profession. The robots that portrayed mechatronics engineers were often regarded craftsmen, while the robots representing gold diggers were often taken as coal miners). This suggests that particularly the robot pairs that represented a dental assistant, a mechatronics engineer, and a gold digger could possibly need further adaptation to optimally portray the respective professions. Nonetheless, participants' guesses about the robots' respective profession were not totally far-fetched. It is thus likely that participants might identify the robots' professions when they are named which was tested in the following.

Table 12.4

The Number of Correct Indications of the Robots' Professions From a First View

Robot Profession	Correct Indications
Construction Worker	35
Perfumery Shop Assistant	35
Soldier	35
Hair Dresser	35
Stewardess	29
Bouncer	27
Butcher	27
Florist	26
Firefighter	25
Cosmetician	25
Housewife	20
Gold Digger	15
Mechatronics Engineer	14
Dental Assistant	0

Note. $N = 36$.

To investigate participants' evaluations of the robots individually, one-sample *t*-tests were performed considering participants' mean ratings of perceived robot gender, the robot professions' recognizability and typicality each relative to the scale midpoint of 4. Ratings of male robot gender should be statistically significantly below the scale midpoint; ratings of female robot gender should be statistically significantly above the scale midpoint. Recognizability and typicality ratings should be at least moderate (indicated by means not statistically significantly different from scale midpoint). Moderate levels were deemed acceptable because the fact that, in Experiment 3, the target's profession and gender were named while participants watched the robots would likely enhance participants' ability to identify the robots' profession and gender. Low levels of recognizability and typicality however would question whether participants could draw a link between visual and verbal input.

According to one-sample *t*-tests that compared participants' mean ratings of robot gender against the scale midpoint of 4, the male robot was clearly perceived as male, while the female robot was clearly perceived as female (see Table 12.5). In line with Bernotat et al. (2017; 2021), the manipulation of the robots' WHR and SW (and hair parts if existing, see Eyssel & Hegel, 2012) evoked the perception of robot gender.

Table 12.5

Results of One-Sample t-Tests Against the Scale Midpoint of 4 on Participants' Evaluations of Robot Gender

Robot Type	Profession	Robot Gender	*M*	*SD*	*t*	*df*	*p*	Cohen's *d*
ASIMO	Bouncer	Male	1.90	1.09	-10.52	29	< .001	1.92
		Female	5.20	1.16	5.68	29	< .001	1.04
	Perfumery Shop Assistant	Male	2.44	1.13	-8.24	35	< .001	1.37
		Female	6.03	0.91	13.37	35	< .001	2.23
Folkwang	Soldier	Male	3.06	1.17	-4.84	35	< .001	0.81
		Female	5.36	1.10	7.43	35	< .001	1.24
	Cosmetician	Male	3.00	1.02	-5.39	29	< .001	0.98
		Female	5.40	0.93	8.23	29	< .001	1.50
Myon	Gold Digger	Male	3.24	1.02	-4.39	33	< .001	0.75
		Female	5.06	0.85	7.26	33	< .001	1.24
	Florist	Male	3.49	0.98	-3.10	34	< .001	0.52
		Female	4.97	0.86	6.71	34	< .001	1.13
NAO	Firefighter	Male	3.40	0.91	-3.88	34	< .001	0.66
		Female	4.54	0.95	3.38	34	< .001	0.57
	Stewardess	Male	3.23	1.03	-4.42	34	< .001	0.75
		Female	4.66	0.91	4.29	34	< .001	0.73

(continued)

Table 12.5

Results of One-Sample t-Tests Against the Scale Midpoint of 4 on Participants' Evaluations of Robot Gender (continued)

Robot Type	Profession	Robot Gender	*M*	*SD*	*t*	*df*	*p*	Cohen's *d*
Ri-Man	Construction Worker	Male	2.56	1.36	-6.37	35	< .001	1.06
		Female	5.58	0.94	10.14	35	< .001	1.69
	Dental Assistant	Male	2.83	1.12	-5.72	29	< .001	1.04
		Female	5.40	1.07	7.17	29	< .001	1.31
Romeo	Mechatronics Engineer	Male	2.50	1.21	-7.21	33	< .001	1.24
		Female	5.09	0.90	7.05	33	< .001	1.21
	Housewife	Male	2.71	0.97	-7.39	30	< .001	1.33
		Female	5.19	1.08	6.17	30	< .001	1.11
SociFlobot	Butcher	Male	3.10	0.88	-5.57	29	< .001	1.02
		Female	5.47	0.90	8.93	29	< .001	1.63
	Hair Dresser	Male	2.91	0.85	-7.53	34	< .001	1.27
		Female	5.54	0.78	11.70	34	< .001	1.98

One-sample *t*-tests were conducted to compare mean ratings of the robot professions' *recognizability* against the scale midpoint. The results show that mean recognizability ratings were at least moderate (see Table 12.6). In terms of the robot professions' recognizability, all robot drawings were thus suitable to be used in Experiment 3. Some minor revisions of the drawings that received only moderate recognizability levels were indicated to depict the robots' professions even more clearly, but not necessary.

Table 12.6

Results of One-Sample t-Tests Against the Scale Midpoint of 4 on Participants' Evaluations of the Robot Professions' Recognizability

Robot Type	Profession	Robot Gender	*M*	*SD*	*t*	*df*	*p*	Cohen's *d*
ASIMO	Bouncer	Male	6.83	0.46	33.65	29	< .001	6.14
		Female	6.63	0.67	21.57	29	< .001	3.94
	Perfumery Shop Assistant	Male	5.67	1.29	7.77	35	< .001	1.29
		Female	5.75	1.20	8.72	35	< .001	1.45
Folkwang	Soldier	Male	6.22	1.17	11.36	35	< .001	1.89
		Female	6.00	1.17	10.25	35	< .001	1.71
	Cosmetician	Male	4.67	1.69	2.16	29	.039	0.39
		Female	4.90	1.45	3.41	29	.002	0.62
Myon	Gold Digger	Male	4.24	1.63	0.84	33	.407	0.14
		Female	3.88	1.79	-0.38	33	.704	0.07
	Florist	Male	3.77	1.68	-0.80	34	.427	0.14
		Female	3.94	1.70	-0.20	34	.843	0.03

(continued)

Table 12.6

Results of One-Sample t-Tests Against the Scale Midpoint of 4 on Participants' Evaluations of the Robot Professions' Recognizability (continued)

Robot Type	Profession	Robot Gender	*M*	*SD*	*t*	*df*	*p*	Cohen's *d*
NAO	Firefighter	Male	3.54	2.02	-1.34	34	.189	0.23
		Female	3.49	1.90	-1.60	34	.119	0.27
	Stewardess	Male	4.69	1.43	2.84	34	.008	0.48
		Female	4.71	1.38	3.05	34	.004	0.52
Ri-Man	Construction Worker	Male	6.67	0.53	29.93	35	< .001	4.99
		Female	6.17	1.00	13.00	35	< .001	2.17
	Dental Assistant	Male	5.80	1.24	7.93	29	< .001	1.45
		Female	6.20	1.00	12.09	29	< .001	2.21
Romeo	Mechatronics Engineer	Male	4.53	1.52	2.03	33	.051	0.35
		Female	4.35	1.41	1.46	33	.154	0.25
	Housewife	Male	4.16	2.19	0.41	30	.685	0.07
		Female	4.55	2.08	1.47	30	.152	0.26
SociFlobot	Butcher	Male	5.23	1.52	4.43	29	< .001	0.81
		Female	5.10	1.27	4.75	29	< .001	0.87
	Hair Dresser	Male	5.49	1.36	6.47	34	< .001	1.09
		Female	5.63	1.29	7.50	34	< .001	1.27

Considering participants' judgements of the robots' professions as being portrayed *typically*, only two typicality ratings of the robots portrayed were statistically significantly below scale midpoint: This concerned the pictures of the female firefighter and of the male florist. The depictions of all other robot professions were judged as moderately to very typical (see Table 12.7)[34]. The robot stimuli were thus suitable for Experiment 3. However, the drawings of the robots as firefighters and florists needed to be revised to optimally portray their assigned profession. In addition, the robots that represented the gold diggers were adapted, too because the profession of a gold digger was rarely encountered nowadays. This might explain why participants could not identify the robot pair as gold diggers when the profession was not explicitly mentioned (see Table 12.4). Therefore, the gold bars and the pickaxe that symbolized the robots' profession on the robot torso were pointed out more clearly. Because the robots' design should be kept constant within-robot pairs, both robots of a pair needed revision (see Appendix A, pp. 286 for the final robot drawings).

[34] The use of multiple *t*-tests should always be regarded with caution. In the present pretest, *t*-tests against the scale midpoint were done to confirm trends in the data reflected by mean values. In most cases, however, the mean values would show a clear tendency even without the comparison against the scale midpoint.

Table 12.7

Results of One-Sample t-Tests Against the Scale Midpoint of 4 on Participants' Evaluations of the Robot Professions' Typicality

Robot Type	Profession	Robot Gender	*M*	*SD*	*t*	*df*	*p*	Cohen's *d*
ASIMO	Bouncer	Male	6.43	1.30	10.22	29	< .001	1.87
		Female	6.03	1.38	8.09	29	< .001	1.48
	Perfumery Shop Assistant	Male	4.14	1.84	0.45	35	.653	0.08
		Female	4.92	1.89	2.91	35	.006	0.49
Folkwang	Soldier	Male	5.64	1.36	7.25	35	< .001	1.21
		Female	5.28	1.50	5.10	35	< .001	0.85
	Cosmetician	Male	3.73	1.84	-0.80	29	.433	0.15
		Female	4.33	1.92	0.95	29	.349	0.17
Myon	Gold Digger	Male	3.82	1.87	-0.55	33	.585	0.09
		Female	3.62	1.91	-1.17	33	.251	0.20
	Florist	Male	3.09	1.93	-2.80	34	.008	0.47
		Female	3.37	1.86	-1.99	34	.054	0.34
NAO	Firefighter	Male	3.37	1.91	-1.95	34	.060	0.33
		Female	2.89	1.62	-4.06	34	< .001	0.69
	Stewardess	Male	4.09	1.72	0.29	34	.770	0.05
		Female	4.17	1.81	0.56	34	.578	0.09
Ri-Man	Construction Worker	Male	6.06	1.07	11.55	35	< .001	1.93
		Female	5.14	1.79	3.81	35	.001	0.64
	Dental Assistant	Male	5.27	1.70	4.08	29	< .001	0.74
		Female	5.53	1.46	5.77	29	< .001	1.05
Romeo	Mechatronics Engineer	Male	4.79	1.61	2.88	33	.007	0.49
		Female	4.26	1.66	0.93	33	.358	0.16
	Housewife	Male	4.00	2.14	< 0.01	30	1.000	< 0.01
		Female	4.29	2.05	0.79	30	.437	0.14
SociFlobot	Butcher	Male	4.43	1.99	1.19	29	.244	0.22
		Female	4.23	1.74	0.74	29	.467	0.13
	Hair Dresser	Male	4.34	1.64	1.23	34	.226	0.21
		Female	4.69	1.66	2.45	34	.020	0.41

Finally, participants' attitudes toward robots as performing a 'profession' in general were explored: When being asked to what extent participants could imagine robots to perform the portrayed professions, means ranged around the scale midpoint, $M = 3.60$, $SD = 1.92$; $t(29) = -1.14$, $p = .264$, $d = 0.21$. This indicates that participants were rather undecided about whether they could imagine robots as performing the presented professions. When they were asked what professions they could imagine robots to perform, they responded they could imagine robots doing simple and dangerous professions that require physical work and no close human-robot interaction, such as factory workers, mechatronics engineers, and soldiers. Participants indicated they could not imagine robots to work in social contexts that would require emotional skills, such as being physicians, teachers, psychologists, or to work in customer service (see also Bernotat & Eyssel, 2018; Bernotat et al.,2021).

12.3 Experiment 3 – Sample

A student sample was recruited at Bielefeld University. Eligibility requirements were the same as in Experiment 1. $N = 87$ participants[4] took part in Experiment 3 (see Section 4.4). 15 of them had to be excluded from data analyses: Eight participants had to be excluded due to technical issues with the eye tracker, two participants had to be excluded because they indicated to be bilingual, and five had to be excluded because they were older than 32 years[10]. The remaining $n = 72$ (male: $n = 35$, female: $n = 37$, $M_{\text{age}} = 24.42$, $SD_{\text{age}} = 3.24$, age range = 19-31 years) were German native speakers with normal or corrected-to normal vision and audition. All participants were naïve about the purpose of Experiment 3 and ensured to have understood the experiment instructions. Furthermore, all individuals were capable of correctly answering the questions followed by the filler trials.

Regarding participants' experience with robots, most participants ($n = 64$) indicated not to have been familiar with one of the displayed robots before. Of those who indicated to have been familiar with one of the robots, only one participant reported to have participated in another study using the NAO robot. Another participant indicated to have been familiar with the robot that portrayed the cosmetician (this is Meka robot which is also used for other studies at CITEC). A third participant reported to have recognized a robot that is commonly used at CITEC, but could not specify which one. Participants who had been familiar with the robots did not have to be excluded from data analyses because prior knowledge of the robot types was unlikely to affect participants' ability to link the robot drawings to the sentences. When being asked from which contexts participants knew robots in general, the majority indicated to know robots from media (media: $n = 46$, books: $n = 28$, work: $n = 22$, home: $n = 19$, other studies: $n = 21$, no experience: $n = 22$, other contexts: $n = 6$).

12.4 Experiment 3 – Design

The experimental design was identical to Experiment 1 (see Section 4.5).

12.5 Experiment 3 – Verbal Stimuli: The Experimental Sentences

The experimental sentences were identical to Experiment 1 (see Section 4.6).

12.6 Experiment 3 – Verbal Stimuli: The Filler Sentences

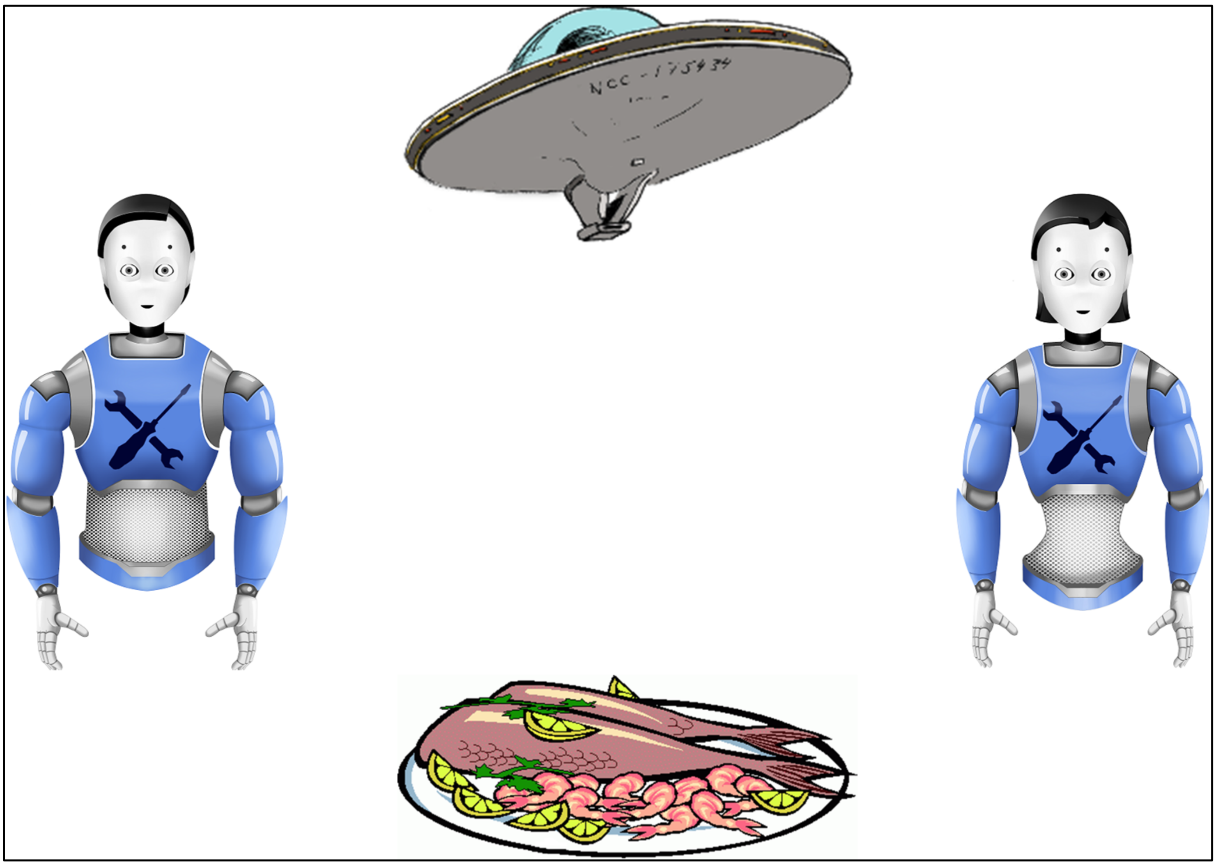
The filler sentences and the related questions were identical to Experiment 1 (see Section 4.7).

12.7 Experiment 3 – Visual Stimuli

The visual scenes that were used in Experiment 3 were similar to Experiment 1 (see Section 4.8). The only difference was that the human characters that were displayed in Experiment 1 were replaced by male and female robotic representatives of the gender-stereotypical professions that were created based on Pretest V (see Section 12.1) and Pretest VI (see Section 12.2; see Appendix A, pp. 223). Figure 12.1 exemplary illustrates the visual scene that referred to Tuple 1 in Table 4.1 (see also Appendix A, Table A17).

Figure 12.1

Visual Scene Referring to Tuple 1 in Table 4.1 Displaying Male and Female Robots



Note. The NP1 object (the fish fillet) is depicted amongst the NP2 characters – a male and a female robot each portrayed as a mechatronics engineer – and an unrelated distractor object.

12.8 Experiment 3 – Questionnaire

To keep experiments comparable, the same questionnaire was used as in Experiment 1 and Experiment 2: Participants' endorsement of benevolent and hostile sexism, normative gender role orientation, internal and external motivation to control for sexist responses, and social desirability were measured (see Section 4.9). Extending the questionnaire used in Experiment 1 and Experiment 2, in Experiment 3 furthermore participants' attitudes toward robots in particular and technology in general as well as their preferences to use robots were assessed. Moreover, participants were asked to indicate to what extent they could imagine robots to perform the presented *tasks* and *professions*, whether the professions were portrayed in a *recognizable* and *typical* manner, and to what extent they perceived one of the robots of a pair as *male* and as *female*. Finally, participants' familiarity with the displayed robots and with robots in general was enquired. These additional measures served to explore participants' attitudes toward robots and technology in general and to explore whether their perceptions of robots in general and of the presented verbal and visual stimuli had affected their visual attention.

Analogous to measures of benevolent and hostile sexism, normative gender role orientation, motivation to control for sexist responses, and social desirability, participants were provided 7-point Likert scales to express their agreement to the statements presented (except for the assessment of participants' preferences for what tasks and profession to use robots, see Section 12.8.8). High values reflected high agreement with the respective statement. Items were recoded if needed. Internal consistencies of the respective constructs (Cronbach's α, see Cronbach, 1951) were calculated per speaker voice condition. Overall, internal consistencies of the measured constructs were moderate to high (Nunnally, 1978[14]), except for IDAQ, which showed only low internal reliabilities in Experiment 3 (see Appendix A, Table B7 for internal consistencies and mean values per speaker voice condition).

The measures reported in the following were used to explore participants' attitudes toward robots and technology in general (see Appendix A, Table A32 – Table A37 for the full set of items):

12.8.1 Robot Anxiety

With eight items (Bernotat et al.,2017;2021), participants' anxiety toward robots was measured (e.g., "I fear one day, robots might take over.").

12.8.2 Robot Acceptance

Ten items (Bernotat & Eyssel, 2018; Eyssel et al.,2011; Ezer, 2008; Reysen, 2005, adapted) assessed to what extent participants would accept having robots in their environment (e.g., "In general, I could imagine to own a robot.").

12.8.3 Agency and Communion

Twelve gender-stereotypical traits were used according to Bem Sex Role Inventory (Bernotat et al.,2017; 2021; Eyssel & Hegel, 2012; Schneider-Düker & Kohler, 1988, adapted) to assess to what extent participants ascribed agentic (e.g., "efficient") and communal traits (e.g., "warmhearted") to robots in general.

12.8.4 Robot Machinelikeness and Humanlikeness

28 items (Haslam, 2006; Loughnan & Haslam, 2007; Schiffhauer, 2015, adapted) served to capture to what extent participants ascribed machinelikeness to robots (e.g., “In general, robots are technical.”) and to what extent they assigned humanlikeness to robots in general (e.g., “In general, robots can understand others’ emotions.”).

12.8.5 Intraindividual Differences in Anthropomorphism (IDAQ; Waytz et al., 2010)

Five items by Waytz et al. (2010) were utilized to capture participants’ proclivity to anthropomorphize non-human entities (e.g., “To what extent does a car have a free will.”).

12.8.6 Technology Commitment (Neyer et al., 2012)

Twelve items by Neyer et al. (2012) measured participants’ self-rated acceptance of technology (e.g., “I am always interested in using new technologies.”), competence in technology use (e.g., “Using new technology is difficult for me. I just cannot handle it.”), and control over technology (e.g., “It’s on me to use technology correctly.”).

12.8.7 Evaluation of the Robot Stimuli and Imaginability of Tasks and Professions

Similar to Pretest VI (see Section 12.2), participants rated to what extent they had *recognized* the robots’ professions, the professions were portrayed in a *typical* way, and whether one of the robots was perceived as *male* and *female.* Moreover, participants stated whether they could imagine robots to perform the *tasks* and *professions* that were mentioned in the sentences.

12.8.8 Tasks and Professions for Which Robots Would be Used

Following Bernotat and Eyssel (2018), using open-response formats, participants named a task for which they would mainly use robots. Then, using 7-point semantic differentials, participants indicated to what extent they deemed the task they had named before as *safe* vs. *dangerous*, *interesting* vs. *boring*, *female* vs. *male*, *socially interactive* vs. *socially isolated*, *demanding* vs. *simple*. High scores reflected high agreement with the latter pole of the differential. Analogously, participants named a profession for which they would mainly use robots and evaluated the profession they had listed in the same manner as they had judged the task.

12.9 Procedure

The procedure was similar to Experiment 1 (see Section 4.10), except that participants were informed that robots would be depicted representing various professions.

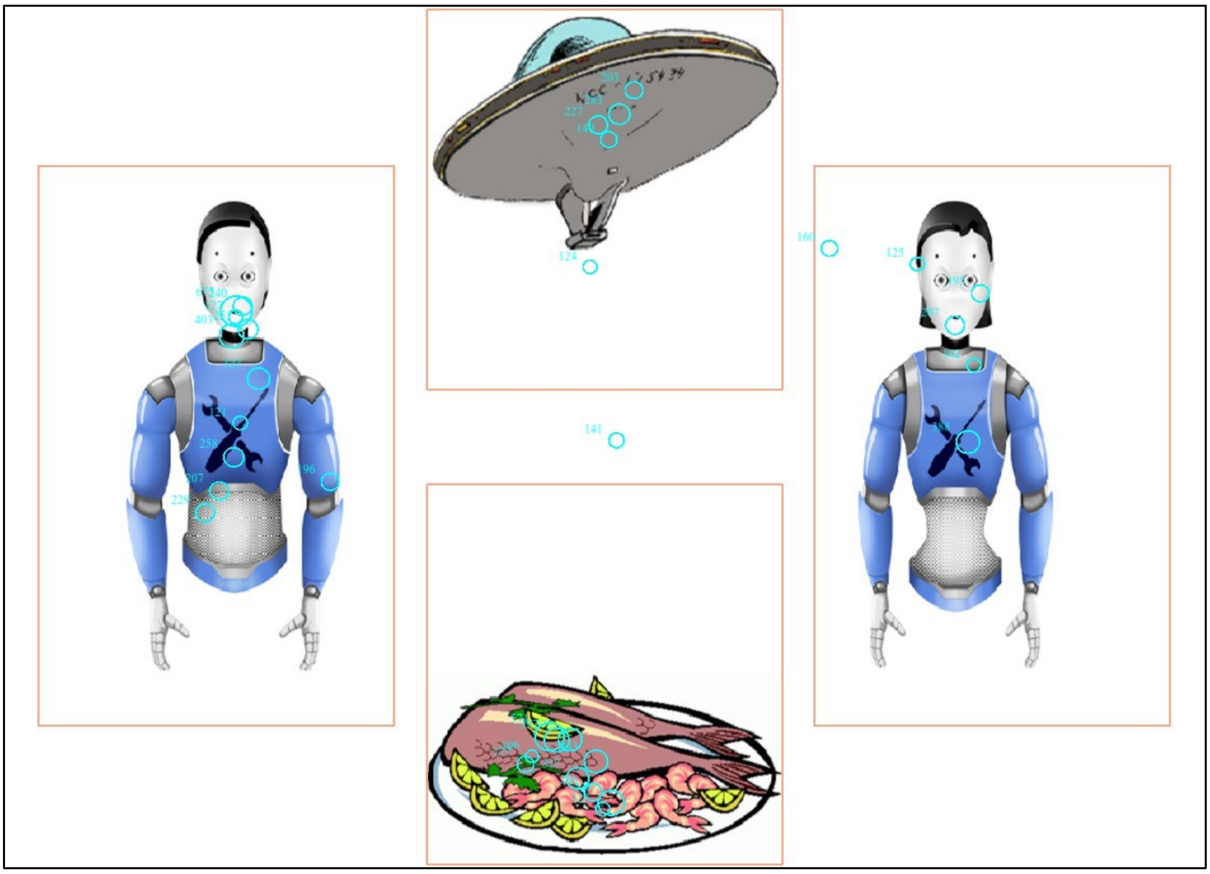
13 Experiment 3 (Adverbs – Robots) – Results

13.1 Experiment 3 – The Steps of Data Processing

In Experiment 3, male and female robots were depicted amongst the NP1 and the distractor object. To capture participants' fixations on the robots, AIs were defined that each framed the male and the female robots (303px x 499px). AIs that framed the objects remained the same as in Experiment 1 (see Section 5.1)[35]. The same experimental sentences were used as in Experiment 1 (see Section 4.6). Therefore, only the AIs that framed the robots needed to be adapted in size (see Figure 13.1). The remaining steps of data processing were done analogously to Experiment 1 (see Section 5.1).

Figure 13.1

The Set of AIs That Framed the Robots and the Objects (Rectangular Frames)[19].



[35] Not to bias participants' visual attention due to stimuli size, the robots and the objects were kept as equal in size as possible (see Holmquist et al., 2011). Nevertheless, the AIs framing the robots had to be slightly longer than those of the objects because a robot's arm length exceeded its torso.

13.2 Experiment 3 – Creating Time Course Graphs

Time course graphs display log-ratios for stereotypically male and female adverbs that either referred to stereotypically male or female NP2 targets (see Figure 13.2 and Figure 13.3). The adverb and NP2 regions' mean on- and offset times were calculated using all fixations across participants and items (see Table 13.1)[36] They were inserted to roughly document word regions.

Table 13.1

Earliest, Latest, and Mean On- and Offset Times (ms) of the Adverb and the NP2 Region

Word Region	Earliest	Latest	*M*	*SD*
Adverb Onset	1913	2647	2215.62	171.88
Adverb Offset	2492	4098	3162.69	309.25
NP2 Onset	4009	6202	5157.57	399.45
NP2 Offset	5036	8637	6356.72	615.04

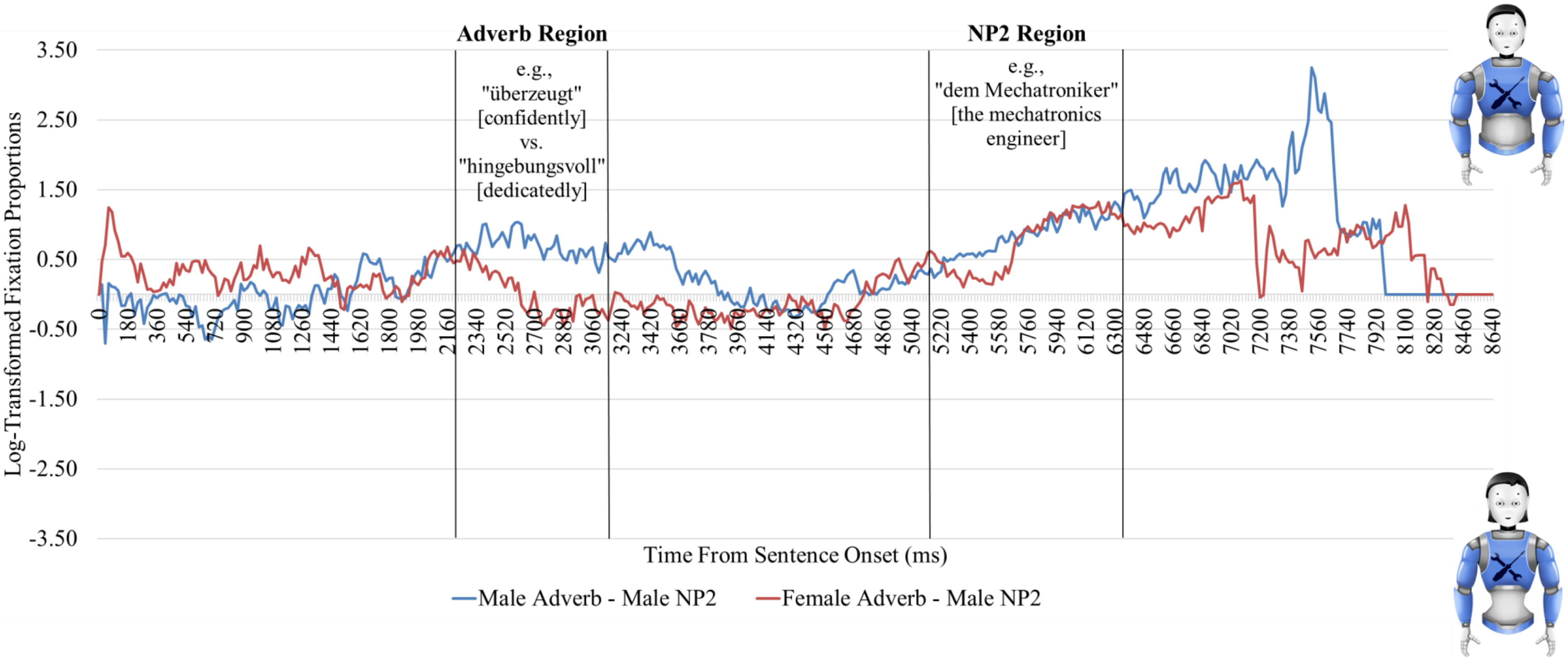
Note. $N = 20863$ (sum of all fixations over participants and items).

[36] In Experiment 3, the same sentences were used as in Experiment 1. The exact word on- and offsets, and durations were thus equal. However, because the number of fixations per participant and per item differed between the two experiments, means and standard deviations of word on- and offset times differed, too.

Figure 13.2 illustrates time course graphs for sentences that ended on a stereotypically male NP2 target: Log-ratios for stereotypically male and female adverbs seemed to diverge shortly after mean adverb onset. This divergence ended around 800ms after mean adverb offset. At a first sight, this might indicate that participants were more likely to look at the target whose gender matched the adverbs' gender-stereotypicality when listening to stereotypically male and female adverbs that referred to a male NP2 target. However, when inspecting log-ratios' range, it becomes obvious that log-ratios for stereotypically male adverbs reached values around 1.0 within the adverb region. Log-ratios for stereotypically female adverbs ranged between around 0.50 and -0.50 within the adverb region. This implies a clear preference for the male robot when the adverbs were stereotypically male, while there was no clear preference for the female robot when the adverbs were stereotypically female. Regarding the NP2 region, the correct NP2 target was anticipated even about 500ms before its mean onset.

Figure 13.2

Time Course Graph for Sentences With a Stereotypically Male vs. Female Adverb Referring to a Stereotypically Male NP2 Target

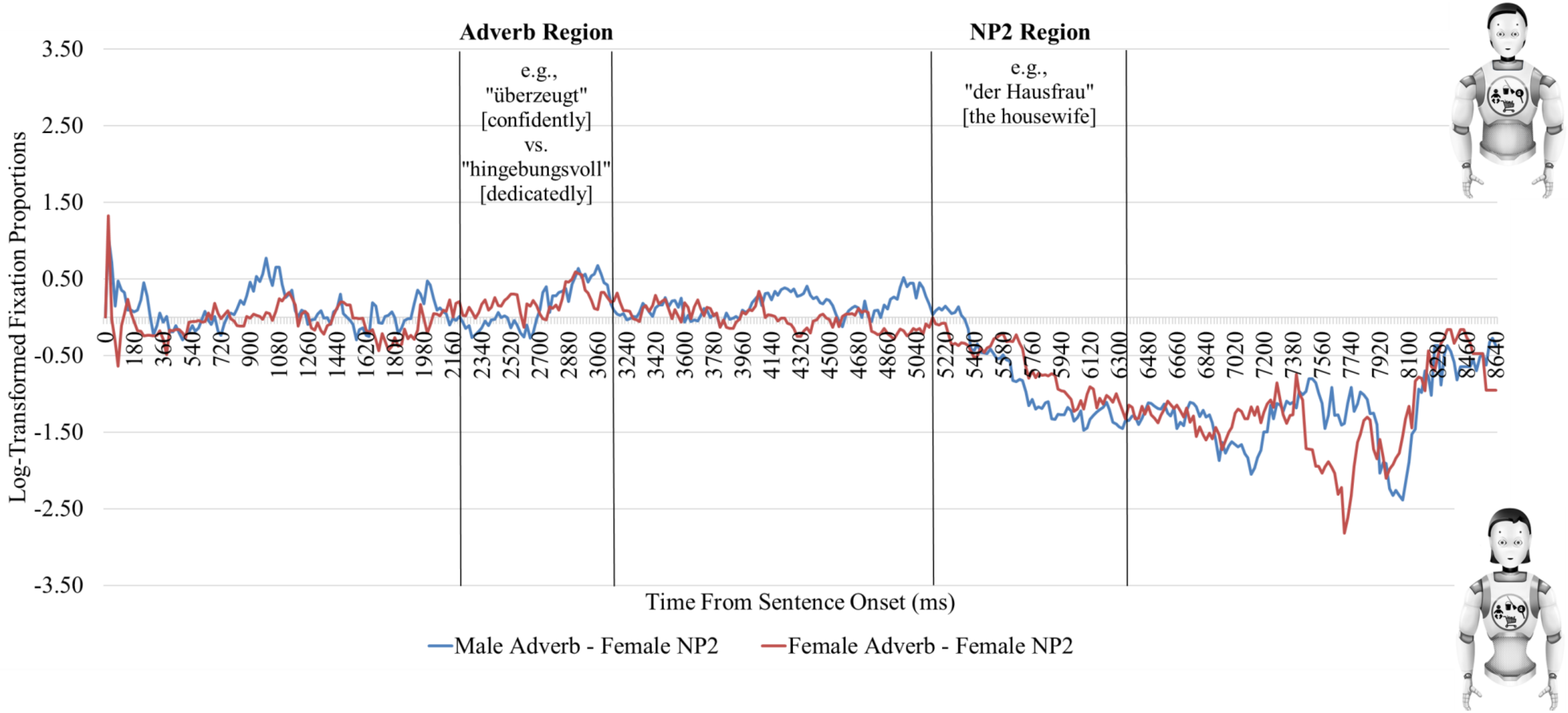


Note. Time course graph depicting changes in log-transformed fixation proportions to the male (positive values) and to the female robot (negative values) when encountering sentences with a stereotypically male (blue line) vs. female (red line) adverb referring to a stereotypically male NP2 target. Mean on- and offset times of the adverb region (1. and 2. vertical line) and mean on- and offset times of the NP2 region (3. and 4. vertical line) are indicated (see Table 13.1).

Figure 13.3 illustrates time course graphs for stereotypically male and female adverbs were followed by a female NP2: No divergence between looks at the male and the female robot occurred within the adverb region. Log-ratios reached rather positive values independent of adverb gender-stereotypicality which indicates that the male robot was generally more frequently fixated than the female one. Regarding the fixation pattern within the NP2 region, negative log-ratios from NP2 mean onset indicate that the correct NP2 target was looked at when it was mentioned. However, inferential statistical analyses are needed to confirm descriptive statistics.

Figure 13.3

Time Course Graph for Sentences With a Stereotypically Male vs. Female Adverb Referring to a Stereotypically Female NP2 Target



Note. Time course graph depicting changes in log-transformed fixation proportions to the male (positive values) and to the female robot (negative values) when encountering sentences with a stereotypically male (blue line) vs. female (red line) adverb referring to a stereotypically female NP2 target. Mean on- and offset times of the adverb region (1. and 2. vertical line) and mean on- and offset times of the NP2 region (3. and 4. vertical line) are indicated (see Table 13.1).

13.3 Experiment 3 – Testing of the Experimental Hypotheses

The adverbs' gender-stereotypicality was hypothesized to lead to fixations on the robot whose gender matched (vs. mismatched) the adverbs' gender-stereotypicality (Hypothesis 1). The gender-matching (vs. mismatching) robot was hypothesized to be more likely fixated when the adverbs were uttered by a speaker whose gender matched (vs. mismatched) the adverbs' gender-stereotypicality (Hypothesis 2). Furthermore, the robot whose gender matched (vs. mismatched) the adverbs' gender-stereotypicality was hypothesized to be more likely fixated, the higher participants' values on benevolent and hostile sexism and normative gender-role orientation (Hypothesis 3a). Participants were expected to less likely fixate the gender-matching robot which would imply more looks at the stereotype-inconsistent robot, the higher their endorsement of internal and external motivation to control for sexist responses and social desirability (Hypothesis 3b).

To test the hypotheses, data were analyzed and reported analogously to Experiment 1 (see Section 5.2; see Appendix B, Table B7 for means on the questionnaire measures; Table B8 for Pearson correlations; Table B27 for post-hoc analyses of statistical power for by-participants and by-items data). Both by-participants data and by-items data were checked for outliers. None of the participants showed any unexpected fixation or response patterns. Likewise, none of the items evoked unexpected response patterns. Requirements for statistical analyses were proofed to be met before inferential analyses were done.

13.3.1 Experiment 3 – The Effects of Adverb Gender-Stereotypicality and Speaker Voice on Log-Ratios by Participants (F_1)

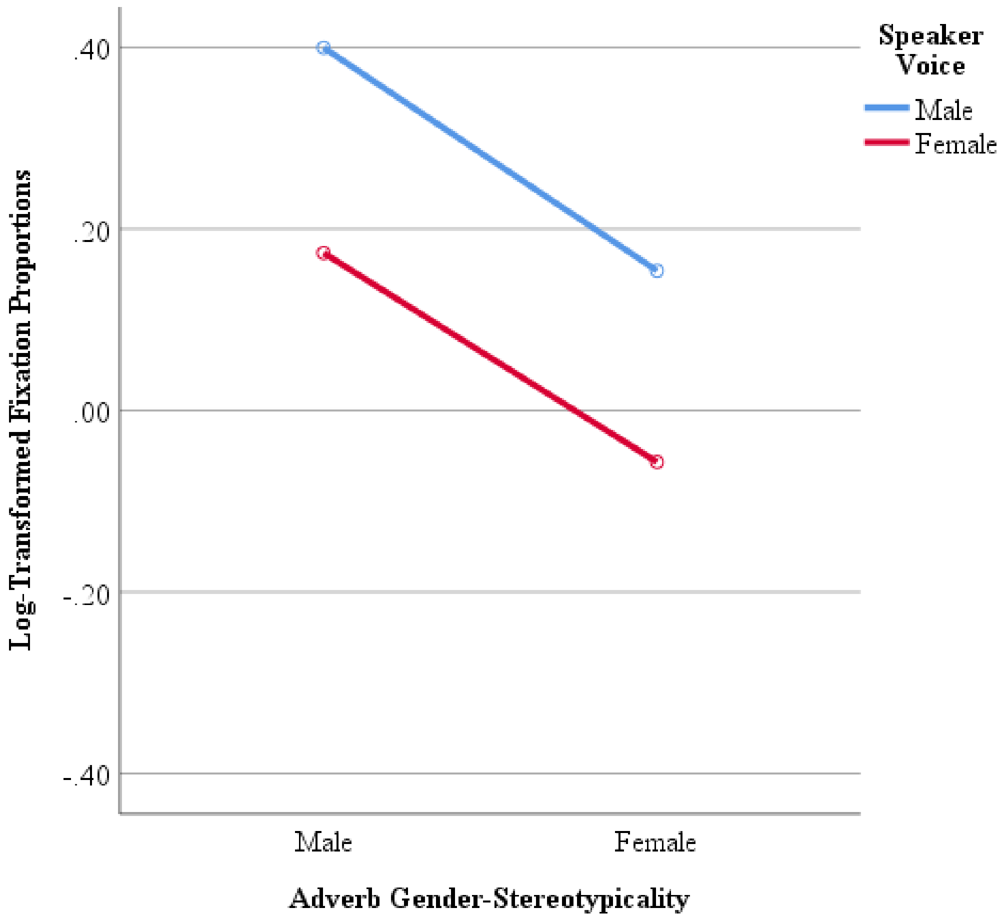
To test Hypothesis 1 and Hypothesis 2 by participants (F_1), a repeated measures MANOVA was conducted with log-ratios calculated by participants as a function of adverb gender-stereotypicality as a within-participants factor and speaker voice as a between-participants factor.

Results of the repeated measures MANOVA revealed no statistically significant main effect of adverb gender-stereotypicality on participants' log-ratios, sphericity assumed: $F_1(1,70) = 2.98$, $p = .089$, $\eta_p^2 = .041$. That is, in analyses conducted by participants, Hypothesis 1 predicting the adverbs' gender-stereotypicality would guide participants' visual attention to a robot whose gender matched the adverbs' gender-stereotypicality was not confirmed. The main effect of speaker voice was not statistically significant, $F_1(1,70) = 3.20$, $p = .078$, $\eta_p^2 = .044$. This indicates that in analyses conducted by participants, Hypothesis 2, predicting a speaker whose gender matched the adverbs' gender-stereotypicality would guide participants' eye gazes to a gender-matching robot was not confirmed, either. Likewise, the interaction between adverb gender-stereotypicality and speaker voice was not statistically significant, sphericity assumed: $F_1(1,70) < 0.01$, $p = .956$, $\eta_p^2 < .001$.

Graphically depicting participants' log-ratios as a function of adverb gender-stereotypicality and speaker voice (Figure 13.4), it becomes apparent that there was a tendency to show the hypothesized fixation pattern: Participants tended to look at the robot whose gender matched the adverbs' gender-stereotypicality, particularly when the adverbs were uttered by a speaker whose gender matched the adverbs' gender-stereotypicality. However, log-ratios ranged between about zero and .40 irrespective of adverb gender-stereotypicality and speaker voice. This indicates that the male robot was generally more likely fixated than the female one.

Figure 13.4

Log-Transformed Fixation Proportions as a Function of Adverb Gender-Stereotypicality and Speaker Voice by Participants (F_1)



13.3.2 Experiment 3 – The Effects of Adverb Gender-Stereotypicality and Speaker Voice on Log-Ratios by Items (F_2)

To test Hypothesis 1 and Hypothesis 2 by items (F_2), a repeated measures MANOVA was conducted with by-items data (F_2). Log-ratios calculated by items were investigated as a function of adverb gender-stereotypicality as a between-items factor and speaker voice as a within-items factor.

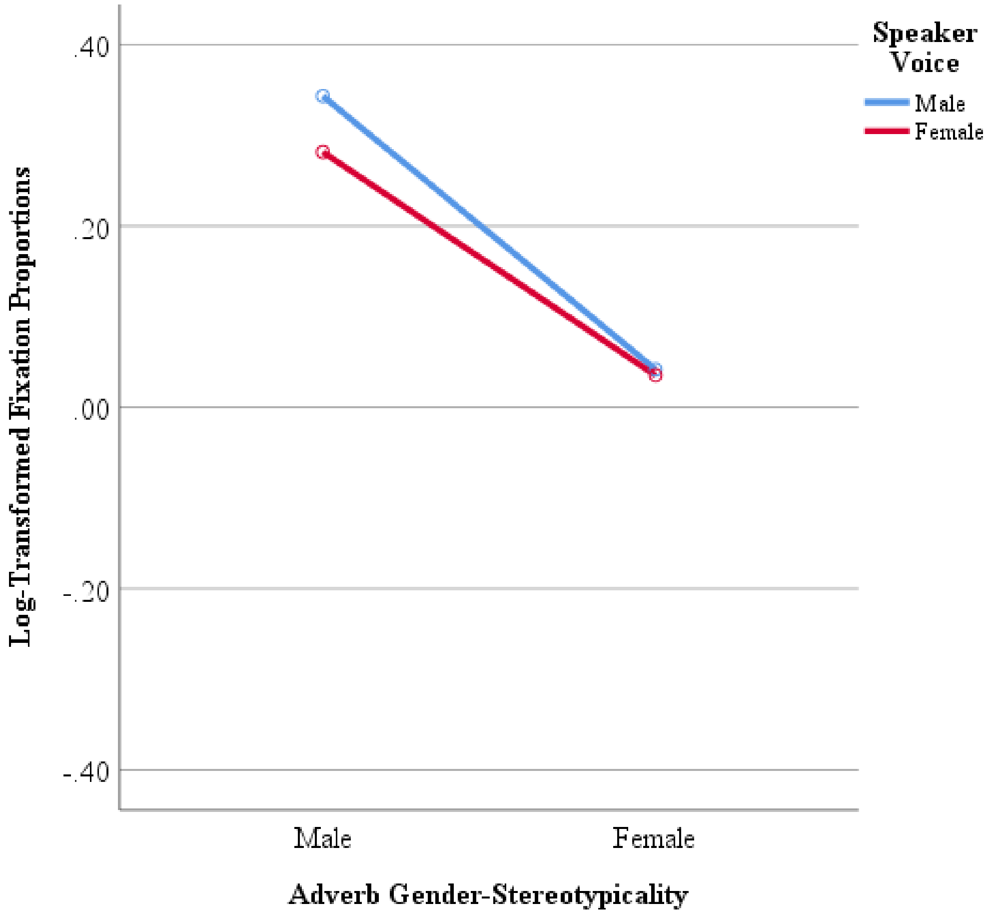
The main effect of adverb gender-stereotypicality was statistically significant, $F_2(1,62) = 4.22$, $p = .044$, $\eta_p^2 = .064$. That is, in line with Hypothesis 1, in by-items analyses, the adverbs' gender-stereotypicality directed participants' eye movements to a robot whose gender matched the adverbs' gender-stereotypicality. The main effect of speaker voice was not statistically significant, sphericity assumed: $F_2(1,62) = 0.09$, $p = .767$, $\eta_p^2 = .001$. Hypothesis 2 predicting the speaker voice would evoke fixations on the robot whose gender matched the adverbs' gender-stereotypicality was not confirmed in by-items analyses. The interaction between adverb gender-stereotypicality and speaker voice was not statistically significant, sphericity assumed: $F_2(1,62) = 0.06$, $p = .810$, $\eta_p^2 = .001$.

Graphically depicting log-ratios as a function of adverb gender-stereotypicality and speaker voice by items (F_2) (see Figure 13.5), it becomes obvious that the items tended evoke the hypothesized fixation pattern: The robot whose gender matched (vs. mismatched) the adverbs' gender-stereotypicality was more frequently looked at, particularly when stereotypically male adverbs were uttered by a male speaker. Overall, however, log-ratios were positive, irrespective of adverb gender-stereotypicality and speaker voice. This implies that the male robot was generally preferred over the female one.

At a descriptive level, the fixation pattern found by items is similar to that found by participants. The fact that the main effect of adverb gender-stereotypicality was statistically significant in by-items analyses, but not in by-participants analyses suggests that variance within-participants might have played a role. Testing Hypothesis 3a and Hypothesis 3b considering the effects of participants' sexist attitudes, normative gender role orientation, motivation to control for sexist responses, and social desirability might explain these findings.

Figure 13.5

Log-Transformed Fixation Proportions as a Function of Adverb Gender-Stereotypicality and Speaker Voice by Items (F_2)



13.3.3 Experiment 3 – The Effects of Adverb Gender-Stereotypicality, Speaker Voice, and Participants' Gender-Related Attitudes on Log-Ratios

To test Hypothesis 3a and Hypothesis 3b, a repeated measures MANCOVA was conducted with by-participants data (F_1). Participants' log-ratios were investigated as a function of adverb gender-stereotypicality and speaker voice. At the same time, participants' endorsement of benevolent and hostile sexism, normative gender role orientation, internal and external motivation to control for sexist responses, and social desirability were considered as covariates.

When the covariates were included, the main effects of adverb gender-stereotypicality, sphericity assumed: $F_1(1,64) = 1.19$, $p = .279$, $\eta_p^2 = .018$, and speaker voice, $F_1(1,64) = 3.36$, $p = .072$, $\eta_p^2 = .050$, and the interaction effect between adverb gender-stereotypicality and speaker voice, sphericity assumed: $F_1(1,64) = 0.09$, $p = .765$, $\eta_p^2 = .001$, were not statistically significant. Regarding the covariates, there was no statistically significant main effect of any of the covariates on participants' log-gazes within the adverb region (see Table 13.2).

Table 13.2

Main Effects of the Covariates on Participants' Log-Ratios Within the Adverb Region

Covariate	F	p	η_p^2
Constant	< 0.01	.984	< .001
External MCSR	2.00	.163	.030
Benevolent Sexism	0.72	.400	.011
Social Desirability	0.16	.690	.002
NGRO	0.11	.740	.002
Hostile Sexism	0.02	.879	< .001
Internal MCSR	0.02	.882	< .001

Note. df (1,64) for all main effects. NGRO = Normative Gender Role Orientation, MCSR = Motivation to Control for Sexist Responses.

However, the interaction effect between adverb gender-stereotypicality and participants' external motivation to control for sexist responses was statistically significant (see Table 13.3). Pearson correlations were calculated to investigate the direction of this interaction effect. Partly in line with Hypothesis 3b, the higher participants' external motivation to control for sexist responses, the more they looked at the female robot when listening to stereotypically male adverbs, $r(70) = -.19$, $p = .114$, and the more they looked at the male robot when listening to stereotypically female adverbs, $r(70) = .28$, $p = .018$.

Table 13.3

Interaction Effects Between Adverb Gender-Stereotypicality and the Covariates

Adverb Gender-Stereotypicality x …	F	p	η_p^2
… External MCSR	7.90	.007	.110
… Social Desirability	3.18	.079	.047
… Hostile Sexism	1.53	.220	.023
… Internal MCSR	0.37	.547	.006
… Benevolent Sexism	0.12	.732	.002
… NGRO	< 0.01	.990	< .001

Note. df(1,64) for all interaction effects. NGRO = Normative Gender Role Orientation, MCSR = Motivation to Control for Sexist Responses.

Following the principle of parsimony (see Tabachick & Fidell, 2007), the MANCOVA was performed again with only external motivation to control for sexist responses as a covariate. The statistically significant interaction effect between adverb gender-stereotypicality and external motivation to control for sexist responses was confirmed, sphericity assumed: $F_1(1,69) = 8.01$, $p = .006$, $\eta_p^2 = .104$. Moreover, the main effect of adverb gender-stereotypicality was statistically significant, sphericity assumed: $F_1(1,69) = 10.56$, $p = .002$, $\eta_p^2 = .133$. The main effect of speaker voice was just not statistically significant, $F_1(1,69) = 3.96$, $p = .051$, $\eta_p^2 = .054$. The main effect of external motivation to control for sexist responses, $F_1(1,69) = 1.34$, $p = .251$, $\eta_p^2 = .019$, and the interaction between adverb gender-stereotypicality and speaker voice, sphericity assumed: $F_1(1,69) = 0.26$, $p = .612$, $\eta_p^2 = .004$, were not statistically significant.

To sum, Hypothesis 3a predicting participants' endorsement of benevolent and hostile sexism and normative gender role orientation would lead to more fixations on the robot whose gender matched (vs. mismatched) the adverbs' gender-stereotypicality was not confirmed. Hypothesis 3b predicting participants' internal and external motivation to control for sexist responses and social desirability would lead to less fixations on the robot whose gender matched (vs. mismatched) the adverbs' gender-stereotypicality, was only partly confirmed for the effects of participants' external motivation to control for sexist responses.

13.4 Experiment 3 – Manipulation Check

Inferential analyses of the NP2 region served as a manipulation check to confirm that the NP2 target was looked at independent of whether it was specified by a male or a female speaker voice. Inferential analyses of the NP2 region were done analogously to Experiment 1.

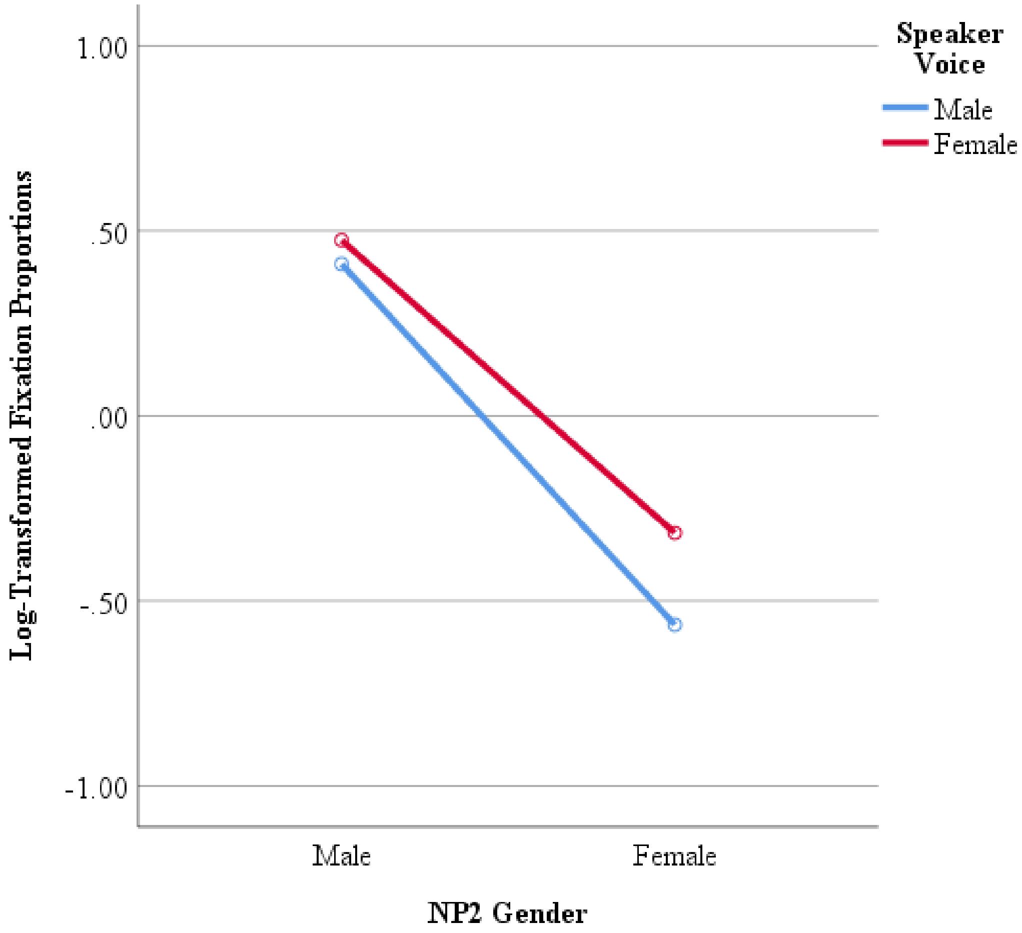
13.4.1 Experiment 3 – The Effects of NP2 Gender and Speaker Voice on Log-Ratios by Participants (F_1)

A repeated measures MANOVA was performed with log-ratios within the NP2 region calculated by participants (F_1) as a function of NP2 gender as a within-participants factor and speaker voice as a between-participants factor.

There was a statistically significant main effect of NP2 gender on participants' eye movements**,** sphericity assumed: $F_1(1,70) = 43.67$, $p < .001$, $\eta_p^2 = .384$. The main effect of speaker voice was not statistically significant, $F_1(1,70) = 2.19$, $p = .144$, $\eta_p^2 = .030$. The interaction between NP2 gender and speaker voice was not statistically significant, either, sphericity assumed: $F_1(1,70) = 0.48$, $p = .493$, $\eta_p^2 = .007$. That is, independent of the speaker voice, participants looked at the correct NP2 target robot when it was named at the end of the sentence. This is also demonstrated in Figure 13.6. Descriptive explorations of participants' fixation pattern moreover demonstrate that a female NP2 target was more likely looked at when it was uttered by a male (vs. female) speaker voice.

Figure 13.6

Log-Transformed Fixation Proportions as a Function of NP2 Gender and Speaker Voice by Participants (F_1)



13.4.2 Experiment 3 – The Effects of NP2 Gender and Speaker Voice on Log-Ratios by Items (F_2)

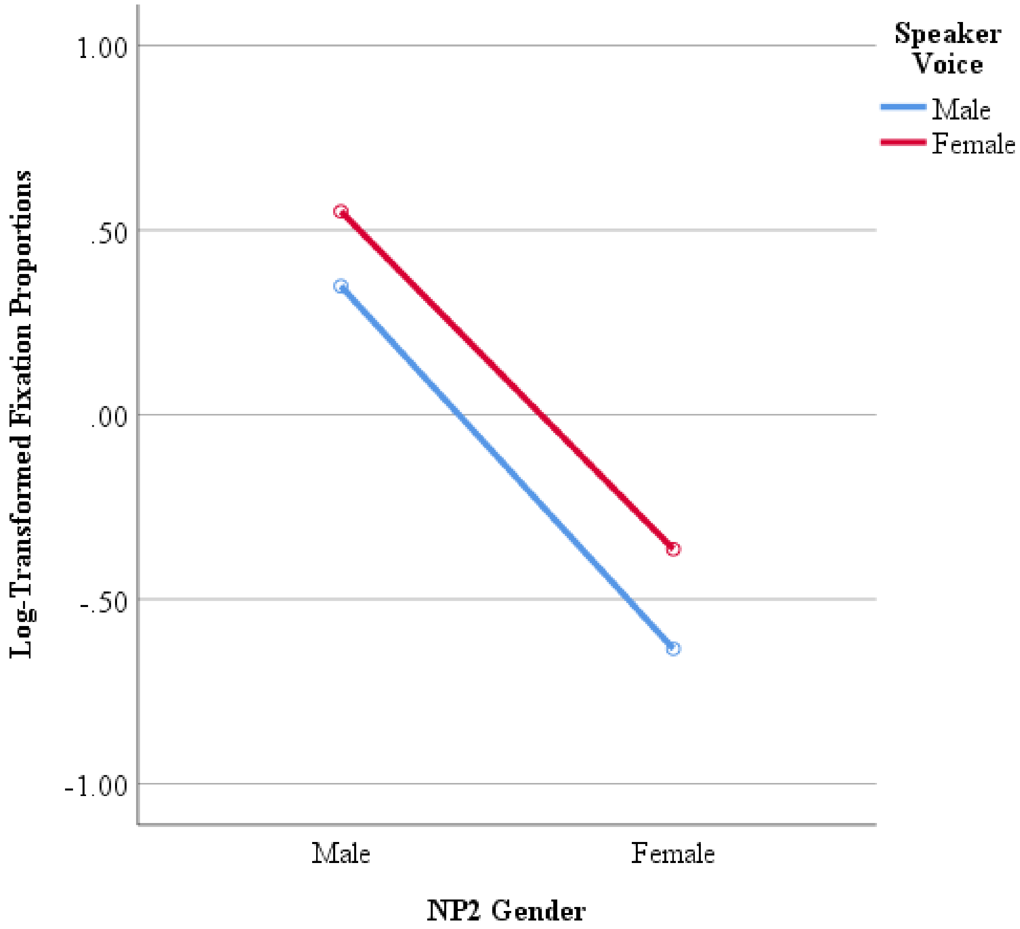
A repeated measures MANOVA was performed with log-ratios within the NP2 region calculated by items (F_2) as a function of NP2 gender as a between-items factor and speaker voice as a within-items factor.

In line with by-participants analyses, the main effect of NP2 gender was statistically significant, sphericity assumed: $F_2(1,62) = 85.90$, $p < .001$, $\eta_p^2 = .581$. That is, the naming of the NP2 target evoked fixations on the correct NP2 target robot. Contrary to results obtained by-participants, the main effect of speaker voice was statistically significant, $F_2(1,62) = 6.12$, $p = .016$, $\eta_p^2 = .090$. That is, moreover, the speaker voice determined whether the male or the female robot was looked at. The interaction between NP2 gender and speaker voice was not statistically significant, sphericity assumed: $F_2(1,62) = 0.13$, $p = .725$, $\eta_p^2 = .002$.

Figure 13.7 illustrates log-ratios by items as a function of NP2 gender and speaker voice: The NP2 target was looked at independent of the speaker voice. However, reflecting the statistically significant main effect of speaker voice, the NP2 target was more likely fixated when it was specified by a speaker whose gender mismatched the NP2 target's gender.

Figure 13.7

Log-Transformed Fixation Proportions as a Function of NP2 Gender and Speaker Voice by Items (F_2)



13.5 Experiment 3 – Exploratory Analyses

Exploratory analyses were conducted to explain the results in more detail. Furthermore, participants' ability to recognize the robots' alleged gender and profession, their ability to imagine robots performing the presented tasks and professions, and their preferences for robot use were explored. Exploratory analyses can be found in Appendix B, pp. 325.

14 Experiment 3 (Adverbs – Robots) – Discussion

This experiment was the third of four visual world experiments. It was an extended replication of Experiment 1. The major aim was to investigate whether the same results would be obtained when gender-stereotypical adverbs referred to robots. The hypotheses, experimental design, and data analyses were thus analogous to Experiment 1. However, unlike Experiment 1, the mentioned professions were portrayed by male and female robots. Moreover, participants' views of robots were explored. In this context, I furthermore tested whether participants' attitudes toward robots, perceptions of the presented robot stimuli, and ability to imagine robots to perform the presented tasks and professions affected their eye movements.

Partly in line with Hypothesis 1, that the adverbs' gender-stereotypicality would guide participants' eye movements to the robot whose gender matched (vs. mismatched) the adverbs' gender-stereotypicality, the main effect of adverb gender-stereotypicality was statistically significant in analyses conducted by items, but not in analyses conducted by participants. Contrary to Hypothesis 2, the main effect of speaker voice was not statistically significant in by-participants and by-items analyses. That means, a speaker whose gender matched (vs. mismatched) the adverbs' gender-stereotypicality did not guide participants' visual attention to the gender-matching robot. The interaction between adverb gender-stereotypicality and speaker voice was not statistically significant in analyses performed by participants and by items.

At a descriptive level, the hypothesized fixation pattern was found: The robot whose gender matched the adverbs' gender-stereotypicality was more likely fixated than the stereotype-inconsistent robot, particularly when the speaker voice matched (vs. mismatched) the adverbs' gender-stereotypicality. Overall, however, independent of adverb gender-stereotypicality and speaker voice, the male robot was more likely fixated than the female one. This fixation pattern supports the assumption based on Experiment 1 that participants could easily comprehend that the adverbs referred to a male or a female target. In Experiment 1, I had assumed this may have been because associations between the adverbs and gender were automatically activated as soon as the adverbs were encountered. Remarkably, in Experiment 1, participants looked at the stereotype-inconsistent character when listening to a male voice. I had deduced that these fixations on the stereotype-inconsistent character might have reflected participants' attempts to counteract stereotypical associations between the adverbs and gender. In the context of Experiment 2, I had concluded that such counteractive fixations may be possible only when language comprehension is gained easily so that participants can be aware of gender stereotypes and have time and cognitive capacities to counteract them (see also Devine, 1989; Fazio, 1990; 2007). Following this rational, the fact that the stereotype-inconsistent robot was not fixated dependent of the speaker voice in Experiment 3 implies that participants may not have been able to counteract the adverbs' gender-stereotypicality because they first needed to comprehend that the sentence's target was represented by a robot. This may have been impeded by their limited ability to imagine robots to perform the presented tasks and professions and by the fact that the professional role names did not explicitly refer to either of the robots. To illustrate, it was referred to "the mechatronics engineer", instead of "the *robot* mechatronics engineer". To nevertheless predict the sentence's target, participants apparently used their knowledge about robots commonly being male (Jung et al.,2016) as the clear preference for the male robot over the female one might suggest. Because the sentence's target always represented a stereotypically male or female profession, participants may have noticed that one of the displayed robots was female. Due to participants' expectations about robots being male, this however may have taken some time. Moreover, due to the facts that the robots were interchangeably displayed on the left or on the right and that they looked very similar because they only differed in WHR and SW, it may have been difficult to discriminate between them (see Wolfe, 1994). After being able to discriminate between the two robots, participants possibly used the speaker voice as an indication of the sentence's target (see Van Berkum et al., 2008). This may have caused the tendency to look at the gender-matching robot when the speaker voice matched the adverbs' gender-stereotypicality. That way, to get acquainted with the presented sentences and the visual scenes may have taken some trials. Then, however, participants may have sensed that the experiment was about gender stereotypes. Similar to Experiment 1 and Experiment 2, participants reported to be rather internally than externally motivated to control for sexist responses. At the same time, they indicated relatively high levels of social desirability. This implies that the idea to possibly behave socially inappropriate or sexist may have evoked feelings of discomfort which they attempted to resolve (see Monteith, 1993; Monteith et al., 2002). Therefore, testing of Hypothesis 3a and Hypothesis 3b, a statistically significant main effect of participants' endorsement of external motivation to control for sexist responses was found. While, contrary to Hypothesis 3a, benevolent and hostile sexism, normative gender role orientation and, contrary to Hypothesis 3b, internal motivation to control for sexist responses and social desirability did not affect participants' eye movements. However, partly in line with Hypothesis 3b, Pearson correlations showed, the higher participants' endorsement of external motivation to control for sexist responses, the more they looked at the stereotype-inconsistent character. Remarkably, when the effect of participants' external motivation to control for sexist responses was controlled for, the main effect of adverb gender-stereotypicality turned out statistically significant. That is, amongst the adverbs' gender-stereotypicality per se, participants' external motivation to control for sexist responses dependent of the adverbs' gender-stereotypicality seemed to have determined their visual attention. This finding strengthens the assumption from Experiment 1, that social, particularly non-sexist norms might have been activated when listening to the sentences. The accessibility of non-sexist norms may have been enhanced by the context of the experiment taking place on a university campus with the experimenter being around (see Blair, 2002; Mitchell et al., 2003). Perhaps participants feared negative consequences not only for showing sexist behavior, but for showing inappropriate social behavior in general when male and female robots were referred to. On the one hand, because gender stereotypes and sexism affect interpersonal relationships between men and women (Glick & Fiske, 1996; 2001), participants may have been uncertain whether they could behave sexist when gendered robots were displayed. On the other hand, they may have been uncertain how others would expect them to behave in the context of robots in general.

This uncertainty in the context of robots is suggested by exploratory analyses. More precisely, it was explored whether participants' evaluations of robots in general and of the presented stimuli had affected their eye movements when listening to the adverbs. The interaction between adverb gender-stereotypicality and participants' ability to imagine robots to perform the presented tasks was statistically significant, while the interaction between adverb gender-stereotypicality and participants' ability to imagine robots to perform the presented professions was just not statistically significant: The more participants indicated to have perceived robots as capable of performing the presented tasks and, in tendency, professions that are commonly done by humans, the more they tended to look at the stereotype-inconsistent robot. At a first sight, these findings may suggest that the more participants perceived robots as humanlike agents that could perform humanlike tasks and roles, the more they aimed to respond non-sexist in the context of robots. However, further exploratory analyses suggest that participants did not necessarily perceive the robots as humanlike. More precisely, the effects of participants' endorsement of external and internal motivation to control for sexist responses and social desirability on their evaluations of robots in general and of the presented robots in particular were explored. Participants' endorsement of external motivation to control for sexist responses had statistically significant effects on their responses on acceptance of technology in general, proclivity to anthropomorphize non-human entities, robot machinelikeness, and robot humanlikeness: The higher participants scored on external motivation to control for sexist responses, the lower levels of robot machinelikeness, the higher levels of robot humanlikeness, proclivity to anthropomorphize non-human entities, and technology acceptance they reported. It seems that the higher participants' endorsement of external motivation to control for sexist responses, the more they might have feared others could consider them as sexist if they would not judge robots as humanlike entities. Due to the fact that gender stereotypes and sexism affect interpersonal relationships (see Glick & Fiske, 2001; Fiske, 1998), participants might have thought they were expected to consider robots as humanlike entities and thus as possible targets of gender-stereotypes and sexism. At the same time, participants' endorsement of social desirability statistically significantly affected their levels of robot acceptance and robot recognizability. That is, the higher participants scored on social desirability, the lower levels of robot acceptance they indicated and the more they reported to have well recognized the robots' professions. Apparently, participants felt socially appropriate not to accept robots, but to have well recognized their designated role. The latter indicates that participants might not have as easily recognized the robots' professions as they had reported. This in turn might support the assumption that participants needed time and cognitive resources to comprehend that the sentence's target was represented by a robot and that one of the displayed robots was female.

Overall, however, in line with participants' responses on the self-report measures, participants apparently succeeded to identify the target's gender and role. This was demonstrated by their fixations at the end of the sentence: Statistically significant main effects of the NP2 gender in by-participants- and by items-analyses indicated that participants looked at the sentence's target as soon as it was explicitly mentioned. According to time course graphs, the NP2 target was looked at even before its mean onset time when it was male. This again suggests that participants' expectations about robots being male may have determined their language processing over the course of the sentence. In tendency, the NP2 target was more likely fixated when it was articulated by a speaker whose gender mismatched its gender. However, only in analyses by items, the main effect of speaker voice was statistically significant. This finding might suggest that participants did not use the speaker voice to identify the sentences' target when it was explicitly specified. Otherwise, the NP2 target would have been more frequently fixated when its gender matched the speaker's gender as within the adverb region. The interaction effect between NP2 gender and speaker voice was not statistically significant in by-participants and by-items analyses. Remarkably, exploratory analyses showed a statistically significant interaction between NP2 gender and participant gender: Female participants looked less frequently at the NP2 target robot than males. Maybe, the robots activated female participants' self-concept of having a lower technology commitment than males. Or, due to their lower robot acceptance, female participants may have less attempted to identify the NP2 target than males.

It seems that participants were generally not enthusiastic about robots. Their wariness of robots may have been due to the fact that most participants reported having little experience with robots in their daily lives. The majority stated to know robots from the media. The image of robots that media create as being supportive on the one hand and threatening on the other may have coined participants' views of robots (see also Bernotat et al.,2021; Bruckenberger et al., 2013; Sandoval et al.,2014). The fact that typical instantiations of robots were displayed may have enhanced the accessibility of this image. Thus unsurprisingly, robots were judged rather as machines than as humanlike agents and rather as agentic than as communal. Though, participants reported a relatively high perceived control over technology and competence in its use, they indicated moderate levels of technology acceptance, robot acceptance, and robot anxiety. As Stapels and Eyssel (2021) pointed out, moderate levels of robot acceptance and anxiety may reflect ambivalence toward robots which may have caused participants' hesitations to interact with them one day. To illustrate, when being asked for what kind of tasks they would use robots for, participants reported to prefer robots for assistive, rather monotonous tasks in which robots' capacities were expected to exceed humans', such as managing huge data sets and transporting goods (see also Bernotat & Eyssel, 2018; Bernotat et al.,2021; Frennert et al.,2013; Horstmann & Krämer, 2019). At the same time, a certain distance should be kept from robots, so that they cannot harm people. These preferences might reflect high trust in robots' functions (cognitive trust), but low trust in their benevolent motives (affective trust; Bernotat et al., 2017; 2021).

To sum, participants' fixation patterns support the assumption that the adverbs were strongly associated with male and female gender. They may further strengthen the conclusion drawn from Experiment 1 and Experiment 2, that during language processing, participants first seemingly attempt to gain language comprehension. When language comprehension is gained and participants are aware of gender stereotypes, they seemingly attempt to counter gender-stereotypes and sexism. This might then lead to counteractive fixations on the stereotype-inconsistent target if cognitive capacities and time are available (see Devine, 1989; Fazio, 1990; 2007). Though the adverbs were apparently effortlessly linked to male or female gender, it seemed to have required additional time and cognitive resources to comprehend that the sentences referred to robots. In line with the Social CIA (Münster & Knoeferle, 2018), participants seemed to use any information available, such as the speaker voice and their knowledge about gender stereotypes and robots for prediction making. Participants seemingly feared disapproval for responding sexist in the context of gendered robots. On the one hand, they seemingly feared being regarded as sexist if they would not consider robots as humanlike entities. On the other hand, they apparently feared disapproval for accepting robots which they, in fact, judged as machinelike tools toward which they might be ambivalent. That way, non-sexist norms and social norms in general might have determined participants' cognitive processes when being exposed to gendered robots.

15 Experiment 4 (Main Verbs - Robots) – Hypotheses

Experiment 4 was an extended replication of Experiment 2. The main difference between the two experiments was that human targets were replaced by robot targets. The hypotheses were thus analogous to Experiment 2:

1. the main verbs' gender-stereotypicality and
2. a speaker voice whose gender matches (vs. mismatches) the main verbs' gender-stereotypicality

lead to higher log-gaze probabilities to look at a robot whose gender matches (vs. mismatches) the main verbs' gender-stereotypicality[1]

Log-gaze probabilities to look at a robot whose gender matches (vs. mismatches) the main verbs' gender-stereotypicality are <u>higher</u>, the higher participants' values on

3a Benevolent and hostile sexism and normative gender role orientation

and <u>lower</u>, the higher participants' values on

3b: Internal and external motivation to control for sexist responses and social desirability.

Analogous to Experiment 3, furthermore participants' perceptions of and attitudes toward robots (i.e., robot human- and machinelikeness, robot agency, robot communion, robot acceptance, robot anxiety), their tendency to anthropomorphize non-human agents, technology commitment, and their evaluations of the presented verbal and visual stimuli (i.e., the robot stimuli's recognizability in terms of gender and profession, the robot professions' typicality, imaginability of robots performing the presented tasks and professions, and robot use preferences) were explored. This served to deepen insights into participants' views of robots in general and of the presented robot stimuli in particular and to moreover examine whether participants' views of robots affected their visual attention.

16 Experiment 4 (Main Verbs - Robots) – Method

16.1 Experiment 4 – Sample

A student sample of $N = 83$ participants[4] was recruited at Bielefeld University. Eligibility requirements were the same as in Experiment 1 (see Section 4.4). Eight participants had to be excluded from data analyses due to technical concerns with the eye tracker or because they turned out to either have been exposed to a second language before the age of six or because they were older than 32 years[10]. As required, the remaining $n = 75$ participants (male: $n = 38$, female: $n = 37$) were German native speakers (German nationality: $n = 71$, other nationality: $n = 4$) not older than 32 years ($M_{\text{age}} = 22.37$, $SD_{\text{age}} = 3.59$, age range: 18-31 years) with normal or corrected-to-normal vision and audition. All respondents were naïve about the purpose of Experiment 4 and ensured to have understood the experiment instructions. Accordingly, all individuals had correctly answered the questions followed by the filler trials.

Regarding participants' experience with robots, most participants ($n = 61$) indicated not to have been familiar with one of the displayed robots. Three of those who indicated to have been familiar with one of the robots reported to have known one of the robots presented from another study (two of them remembered that they had used the NAO robot in a prior study). Participants who had been familiar with the robots did not have to be excluded from data analyses because prior knowledge of the robot types was unlikely to affect their ability to link the robot drawings to the sentences. When being asked from which contexts participants knew robots in general, most indicated to know robots from media (media: $n = 47$, books: $n = 23$, work: $n = 15$, home: $n = 11$, other studies: $n = 6$, no experience: $n = 15$, other contexts: $n = 6$, multiple indications possible).

16.2 Experiment 4 - Design

The experimental design was similar to Experiment 2 (see Section 8.3).

16.3 Experiment 4 – Verbal Stimuli: Experimental Sentences Including Gender-Stereotypical Main Verbs

The experimental sentences were identical to Experiment 2 (see Section 8.4).

16.4 Experiment 4 – Verbal Stimuli: Filler Sentences

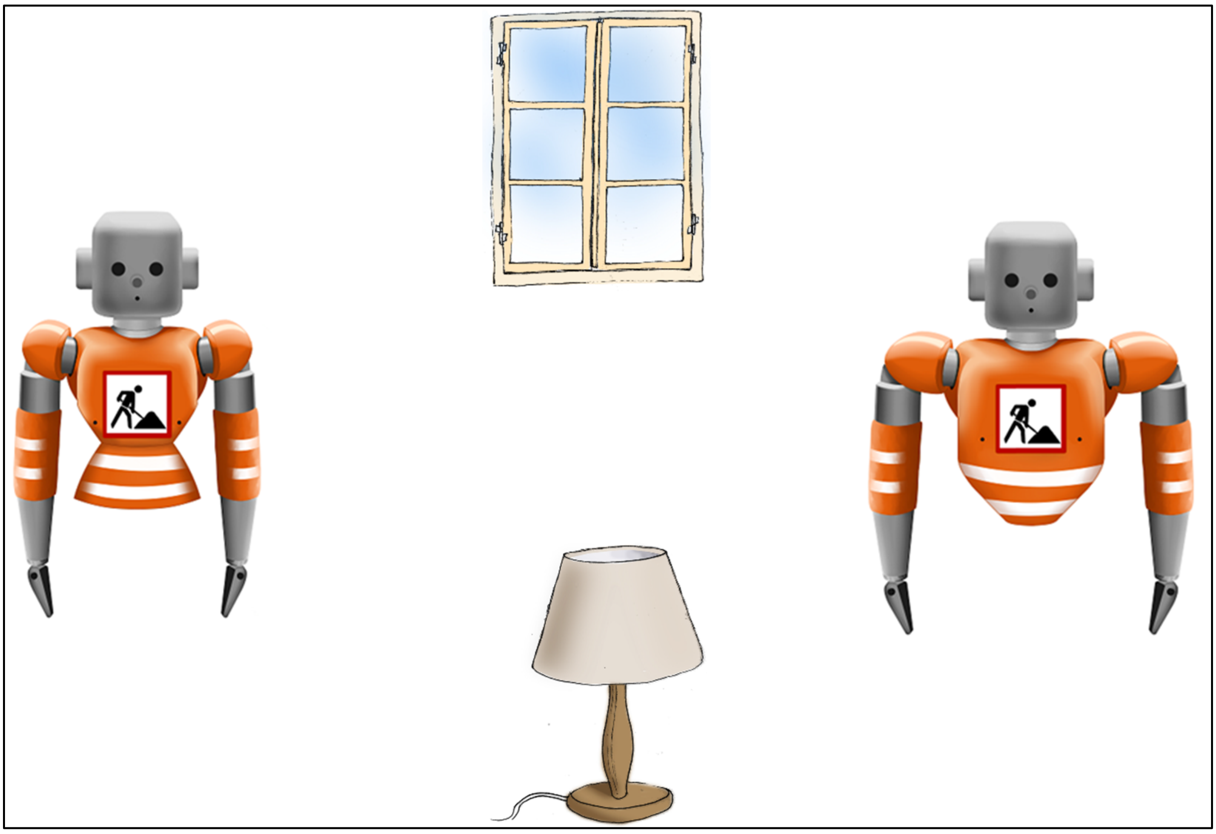
The filler sentences and the related questions were identical to Experiment 2 (see Section 8.5).

16.5 Experiment 4 – Visual Stimuli

The visual scenes were similar to Experiment 2 (see Section 8.6). The only difference was that the human characters that were displayed in Experiment 1 were replaced by male and female robotic representatives of the gender-stereotypical professions. Figure 16.1 exemplary illustrates the visual scene that referred to Tuple 1 in Table 8.1 (see also Appendix A, Table A and Table A29 for the full set of experimental sentences and filler sentences).

Figure 16.1

Visual Scene Referring to Tuple 1 in Table 8.1 Displaying Male and Female Robots



Note. The NP1 object (the window) is depicted amongst the NP2 characters – a male and a female robot each portrayed as a construction worker – and an unrelated distractor object.

16.6 Experiment 4 – Questionnaire

The questionnaire was identical to Experiment 3 (see Section 12.8). For most of the constructs, internal consistencies of the measured constructs were moderate to high (see Nunnally, 1978[14]). Only IDAQ showed low internal reliabilities (see Appendix A, Table B17 for internal consistencies and mean values per speaker voice condition).

16.7 Experiment 4 – Procedure

The procedure was identical to Experiment 3 (see Section 12.9).

17 Experiment 4 (Main Verbs - Robots) – Results

17.1 Experiment 4 – The Steps of Data Processing

In Experiment 4, the same set of AIs were used as in Experiment 3 (see Figure 13.1) to capture fixations on the male and the female robot. Because the same experimental sentences were used as in Experiment 2, the steps of data processing were done analogously to Experiment 2 (see Section 9.1).

17.2 Experiment 4 – Creating Time Course Graphs

Using log-ratios for stereotypically male and female main verbs that either referred to a stereotypically male or female NP2 target, participants' time-lined fixation patterns to the male and the female robot were depicted over the course of the sentence (see Figure 17.1 and Figure 17.2). The main verb and NP2 regions' mean on- and offset times (see Table 17.1)[37] were calculated using all fixations across participants and items. They served to roughly document word regions.

Table 17.1

Earliest, Latest, and Mean On- and Offset Times (ms) of the Main Verb and the NP2 Region

Word Region	Earliest	Latest	*M*	*SD*
Main Verb Onset	1172	2363	1860.84	221.98
Main Verb Offset	1903	3363	2691.33	280.43
NP2 Onset	3258	4440	3814.28	287.05
NP2 Offset	4252	6615	5079.28	464.22

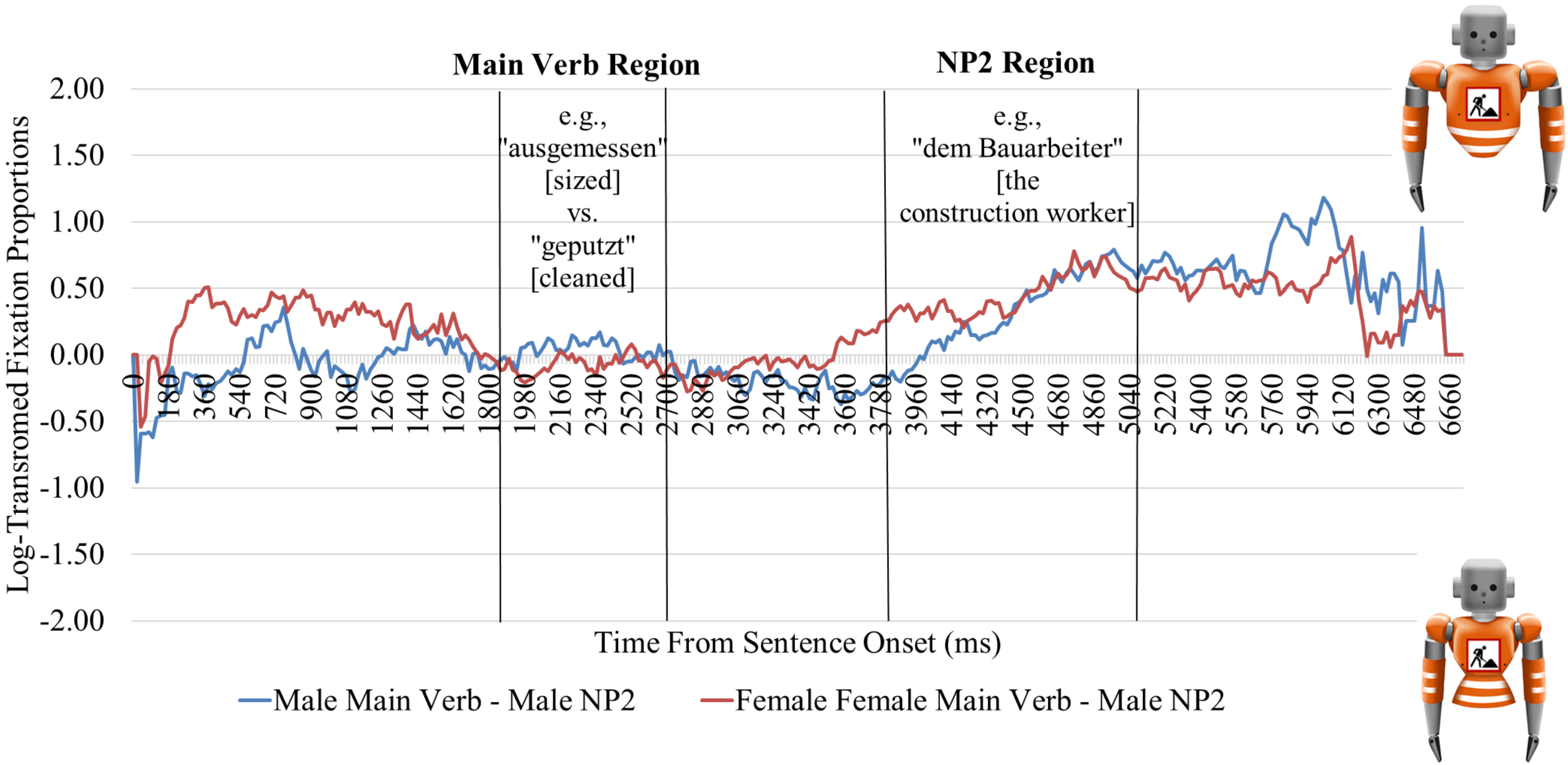
Note. N = 16317 (sum of all fixations over participants and items).

[37] The same experimental sentences were used as in Experiment 2. Therefore, the exact on- and offsets, and duration times were the same as in Experiment 2. However, mean on-and offsets differed across experiments because they depended on the number of fixations an item attracted per participant. Furthermore, as in Experiment 2, tuples 7, 22, and 28 had to be excluded from data analyses due to inconsistencies.

Figure 17.1 illustrates time course graphs for sentences ending on a stereotypically male NP2 target: Log-ratios within the main region ranged around zero, independent of the main verbs' gender-stereotypicality. This implies that the male and the female robot were fixated equally likely within the main verb region. Regarding log-ratios within the NP2 region, the male NP2 target was looked at from mean NP2 onset and even shortly before when stereotypically female main verbs anteceded the male NP2 target. Log-ratios within the NP2 region reached values up to about 1.25 around mean NP2 offset which indicates a clear preference for the male NP2 target when it was mentioned at the end of the sentence.

Figure 17.1

Time Course Graph for Sentences With a Stereotypically Male vs. Female Main Verb Referring to a Stereotypically Male NP2 Target

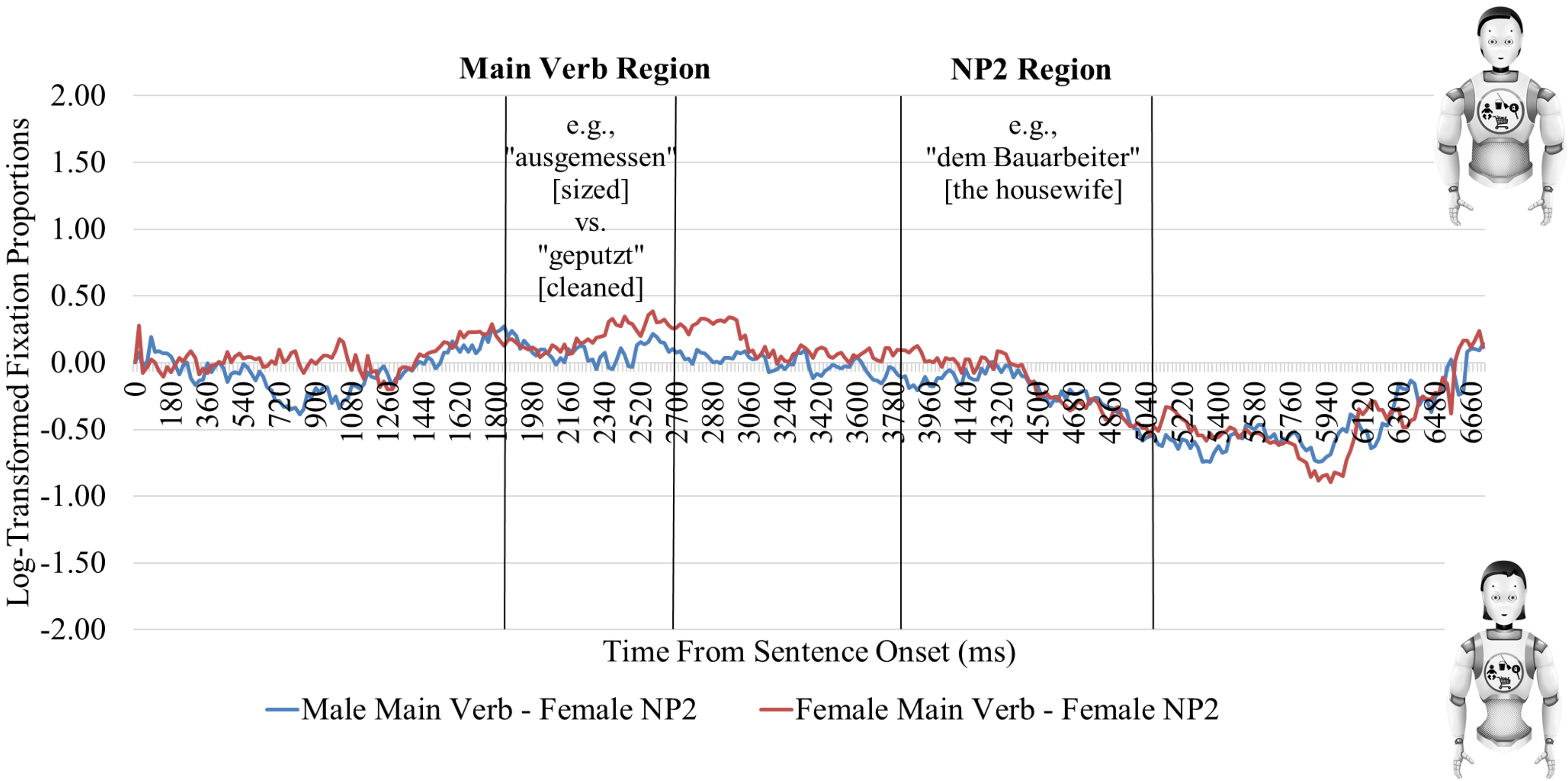


Note. Time course graph depicting changes in log-transformed fixation proportions to the male (positive values) and to the female robot (negative values) when encountering sentences with a stereotypically male (blue line) vs. female (red line) main verb referring to a stereotypically male NP2 target. Mean on- and offset times of the main verb region (1. and 2. vertical line) and mean on- and offset times of the NP2 region (3. and 4. vertical line) are indicated (see Table 17.1).

Figure 17.2 visualizes time course graphs for sentences ending on a female NP2 target: Log-ratios within the main verb region were positive with values below 0.50, irrespective of whether the main verbs were stereotypically male or female. This indicates that there was a tendency to look at the male robot, irrespective of the main verbs' gender-stereotypicality. Regarding fixation patterns within the NP2 region, the NP2 target was looked at about 500ms after mean NP2 onset. Log-ratios reached values up to about -0.50 at mean NP2 offset. This implies that, compared to log-ratios within the NP2 region for sentences ending on a male NP2 target, the female NP2 target robot was later and less likely fixated than the male one.

Figure 17.2

Time Course Graph for Sentences With a Stereotypically Male vs. Female Main Verb Referring to a Stereotypically Female NP2 Target



Note. Time course graph depicting changes in log-transformed fixation proportions to the male (positive values) and to the female robot (negative values) when encountering sentences with a stereotypically male (blue line) vs. female (red line) main verb referring to a stereotypically female NP2 target. Mean on- and offset times of the main verb region (1. and 2. vertical line) and mean on- and offset times of the NP2 region (3. and 4. vertical line) are indicated (see Table 17.1).

17.3 Experiment 4 – Testing of the Experimental Hypotheses

The main verbs' gender-stereotypicality was hypothesized to lead to fixations on the robot whose gender matched (vs. mismatched) the main verbs' gender-stereotypicality (Hypothesis 1). The gender-matching (vs. mismatching) robot was hypothesized to be more likely fixated when the main verbs were uttered by a speaker whose gender matched (vs. mismatched) their gender-stereotypicality (Hypothesis 2). Furthermore, the gender-matching robot was hypothesized to be more likely fixated, the higher participants' values on benevolent and hostile sexism and normative gender-role orientation (Hypothesis 3a). Participants were expected to less likely fixate the gender-matching robot which would imply more looks at the stereotype-inconsistent robot, the higher their endorsement of internal and external motivation to control for sexist responses and social desirability (Hypothesis 3b). To test the hypotheses, data were analyzed and reported analogously to Experiment 2 (see Section 9.2; see Appendix B, Table B17 for means on the questionnaire measures; Table B18 for Pearson correlations; and Table B27 for post-hoc analyses of statistical power for by-participants and by-items data).

Data processed by participants (including questionnaire data) and by items were examined for outliers. No outliers were identified. Consequently, none of the participants showed unexpected fixation patterns nor unexpected response patterns on the questionnaire measures. Likewise, none of the items evoked unexpected fixation patterns. Requirements for statistical analyses were proofed to be met before inferential analyses were done.

17.3.1 Experiment 4 – The Effects of Main Verb Gender-Stereotypicality and Speaker Voice on Log-Ratios by Participants (F_1)

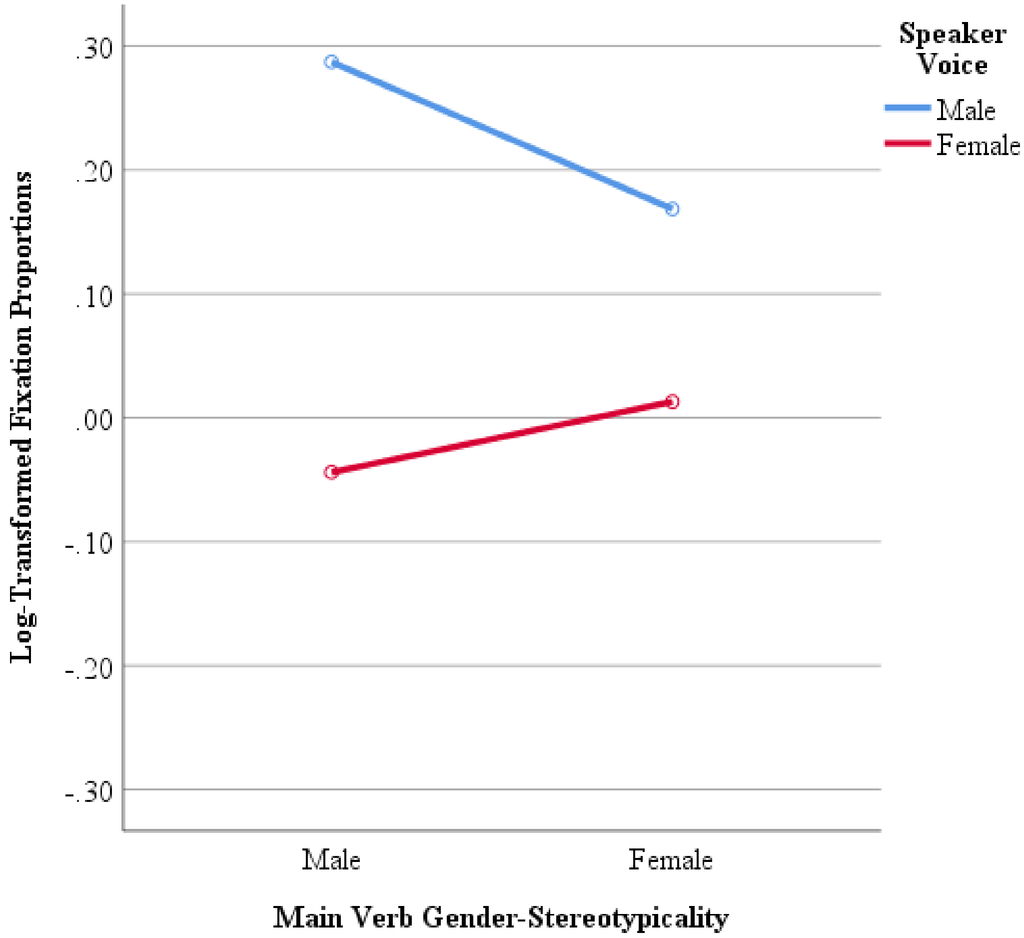
To test the Hypothesis 1 and Hypothesis 2, a repeated measures MANOVA was conducted with log-ratios calculated by participants as a function of main verb gender-stereotypicality as a within-participants factor and speaker voice as a between-participants factor.

The main effects of main verb gender-stereotypicality, sphericity assumed: $F_1(1,69) = 0.34$, $p = .854$, $\eta_p^2 < .001$, and speaker voice, $F_1(1,69) = 3.11$, $p = .082$, $\eta_p^2 = .043$, and the interaction between main verb gender-stereotypicality and speaker voice, sphericity assumed: $F_1(1,69) = 0.28$, $p = .599$, $\eta_p^2 = .004$, were not statistically significant. That is, contrary to Hypothesis 1 and Hypothesis 2, in analyses conducted by participants, the main verbs' gender-stereotypicality and the speaker voice did not guide participants' eye gazes to a target whose gender matched the main verbs' gender-stereotypicality.

Figure 17.3 illustrates participants' log-ratios as a function of main verb gender-stereotypicality and speaker voice: When listening to a male speaker, participants tended to look at the robot whose gender matched the adverbs' gender-stereotypicality. Overall, however, log-ratios when listening to the male speaker were positive. This indicates that the male robot was clearly preferred over the female one when listening to a male speaker, irrespective of the main verbs' gender-stereotypicality. When listening to a female speaker, there was a slight tendency to look at the stereotype-inconsistent robot. Log-ratios ranged around -0.05 for stereotypically male main verbs and around zero for stereotypically female main verbs. This indicates that both robots were fixated equally likely with a slight tendency to look at the female robot when stereotypically male main verbs were uttered by a female speaker voice.

Figure 17.3

Log-Transformed Fixation Proportions as a Function of Main Verb Gender-Stereotypicality and Speaker Voice by Participants (F_1)



17.3.2 Experiment 4 – The Effects of Main Verb Gender-Stereotypicality and Speaker Voice on Log-Ratios by Items (F_2)

To test Hypothesis 1 and Hypothesis 2 by items, a repeated measures MANOVA was computed with log-ratios calculated by items as a function of main verb gender-stereotypicality as a between-items factor and with speaker voice as a within-items factor.

The main effect of main verb gender-stereotypicality, $F_2(1,56) = 0.25$, $p = .618$, $\eta_p^2 = .004$, was not statistically significant. That is, contrary to Hypothesis 1, in analyses conducted by items, the main verbs' gender-stereotypicality did not guide participants' eye gazes to a robot whose gender matched the main verbs' gender-stereotypicality.

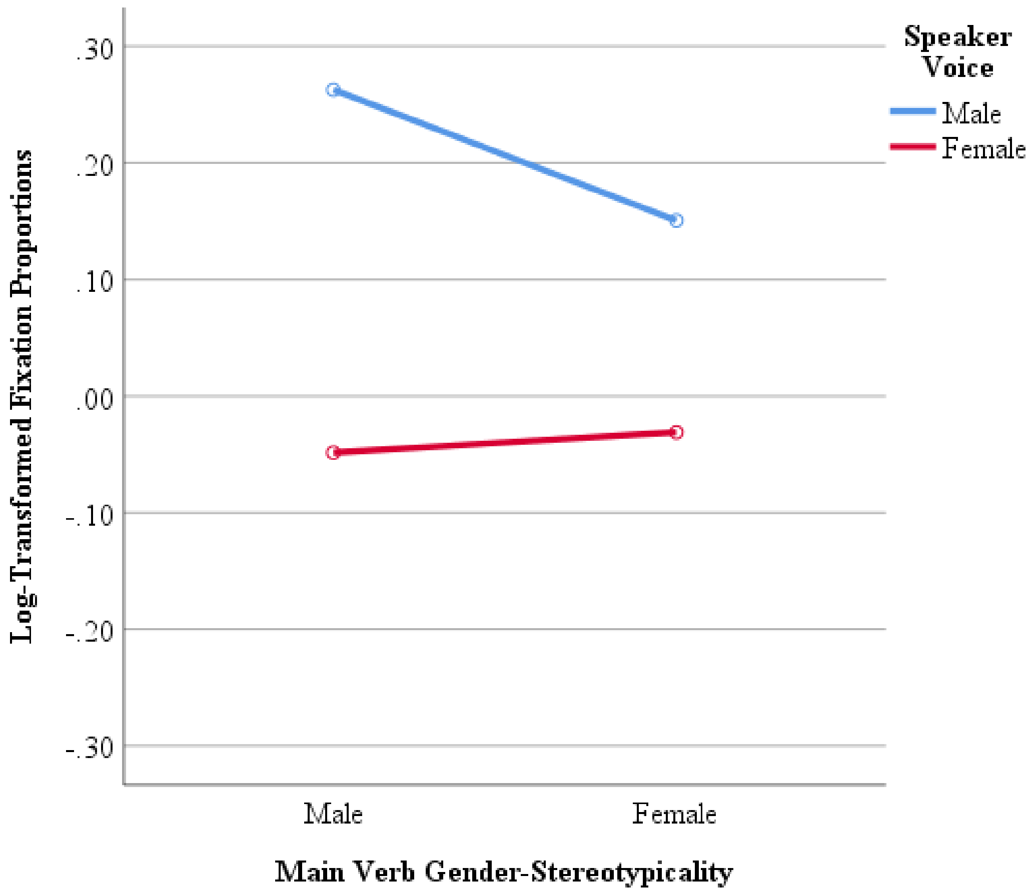
However, the main effect of speaker voice was statistically significant, sphericity assumed: $F_2(1,56) = 6.06$, $p = .017$, $\eta_p^2 = .098$. That is, in analyses conducted by items, the speaker voice determined whether the male or the female robot was looked at. The interaction between main verb gender-stereotypicality and speaker voice was not statistically significant, sphericity assumed: $F_2(1,56) = 0.42$, $p = .520$, $\eta_p^2 = .007$.

Figure 17.4 displays log-ratios by items as a function of main verb gender-stereotypicality and speaker voice: Similar to the fixation pattern found by participants, there was a tendency to look at the robot whose gender matched the adverbs' gender-stereotypicality when the main verbs were uttered by a male speaker. However, log-ratios when listening to a male speaker were positive, irrespective of the main verbs' gender-stereotypicality. This indicates a clear preference to look at the male robot when listening to a male speaker. Log-ratios when listening to the female speaker voice were negative with values about -0.05, independent of the main verbs' gender-stereotypicality. This indicates a slight preference to look at the female robot when listening to a female speaker. That is, reflecting the main effect of speaker voice, the robot whose gender matched (vs. mismatched) the speaker's gender was more likely looked at, independent of the main verbs' gender-stereotypicality. This pattern was however more apparent when listening to a male (vs. female) speaker voice. That is, although the main effect of speaker voice was statistically significant in by-items analyses, Hypothesis 2 predicting the speaker voice would guide participants' visual attention to the robot whose gender matched the main verbs' gender-stereotypicality was not confirmed.

The fact that the main effect of speaker voice was statistically significant in by-items analyses, but not in by-participants analyses suggests that variance across participants might have affected participants' fixation patterns when listening to the main verbs. Testing for the effects of participants' endorsement of benevolent and hostile sexism, normative gender role orientation, internal and external motivation to control for sexist responses, and social desirability in the context of Hypothesis 3 might therefore explain these findings in more detail.

Figure 17.4

Log-Transformed Fixation Proportions as a Function of Main Verb Gender-Stereotypicality and Speaker Voice by Items (F_2)



17.3.3 Experiment 4 – The Effects of Main Verb Gender-Stereotypicality, Speaker Voice, and Participants' Gender-Related Attitudes on Log-Ratios

To test Hypothesis 3a and Hypothesis 3b, a repeated measures MANCOVA was conducted with log-ratios by participants as a function of main verb gender-stereotypicality as a within-participants factor and speaker voice as a between-participants factor. The effects of participants' scores on benevolent and hostile sexism, normative gender role orientation, external and internal motivation to control for sexist responses, and social desirability were considered as covariates.

In line with by-participants results when the covariates were not considered, the main effects of main verb gender-stereotypicality, sphericity assumed: $F_1(1,63) = 0.09$, $p = .768$, $\eta_p^2 = .001$, and speaker voice, $F_1(1,63) = 3.23$, $p = .077$, $\eta_p^2 = .049$, and the interaction between main verb gender-stereotypicality and speaker, sphericity assumed: $F_1(1,63) = 0.49$, $p = .486$, $\eta_p^2 = .008$, voice were not statistically significant. Likewise, the main effects of the covariates were not statistically significant (see Table 17.2).

Table 17.2

Main Effects of the Covariates on Participants' Log-Ratios Within the Main Verb Region

Covariate	*F*	*p*	η_p^2
NGRO	.86	.356	.014
Social Desirability	.81	.372	.013
Internal MCSR	.15	.704	.002
Benevolent Sexism	.13	.720	.002
External MCSR	.05	.818	.001
Hostile Sexism	.03	.863	< .001

Note. *df*(1,63) for all main effects. NGRO = Normative Gender Role Orientation, MCSR = Motivation to Control for Sexist Responses.

Likewise, none of the interaction effects between main verb gender-stereotypicality and the covariates was statistically significant (see Table 17.3).

Table 17.3

Interaction Effects Between Main Verb Gender-Stereotypicality and the Covariates on Participants' Log-Ratios Within the Main Verb Region

Main Verb Gender-Stereotypicality x Covariate	*F*	*p*	η_p^2
… Benevolent Sexism	.589	.446	.009
… NGRO	.398	.530	.006
… Hostile Sexism	.199	.657	.003
… Social Desirability	.056	.813	.001
… Internal MCSR	.033	.856	.001
… External MCSR	.004	.953	.000

Note. *df*(1,63) for all interaction effects. NGRO = Normative Gender Role Orientation, MCSR = Motivation to Control for Sexist Responses.

That is, neither the main verbs' gender-stereotypicality and the speaker voice nor any of the covariates determined whether the male or the female robot was looked at (see Table B17 for participants' mean scores on the covariates; see Table B18 for Pearson correlations; see Table 27 for post-hoc analyses of statistical power by participants and by items).

17.4 Experiment 4 – Manipulation Check

Inferential analyses of the NP2 region served as a manipulation check to confirm that the NP2 target was looked at independent of whether it was specified by a male or a female speaker voice. Inferential analyses of the NP2 region were done analogously to Experiment 1.

17.4.1 Experiment 4 – The Effects of NP2 Gender and Speaker Voice on Log-Ratios by Participants (F_1)

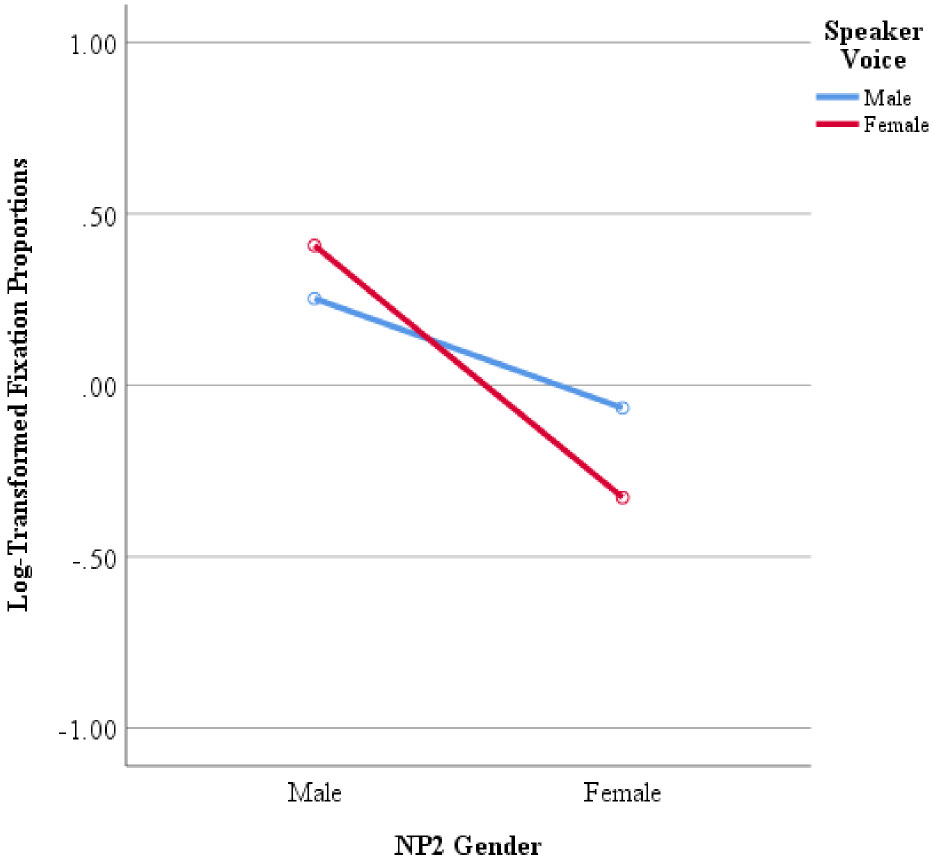
A repeated measures MANOVA was computed with log-ratios calculated by participants as a function of NP2 gender as a within-participants factor and speaker voice as a between-participants factor.

The main effect of NP2 gender was statistically significant, sphericity assumed, $F_1(1,72) = 25.67$, $p < .001$, $\eta_p^2 = .263$. That is, NP2 gender determined whether the male or the female robot was looked at. The main effect of speaker voice, $F_1(1,72) = 0.38$, $p = .541$, $\eta_p^2 = .005$, was not statistically significant. The interaction between NP2 gender and speaker voice was just not statistically significant, sphericity assumed, $F_1(1,72) = 3.99$, $p = .050$, $\eta_p^2 = .053$.

Figure 17.5 displays participants' log-ratios by participants as a function of NP2 gender and speaker voice: As indicated by the statistically significant main effect of NP2 gender, the NP2 target was looked when its gender was specified. This was independent of the speaker voice. However, at a descriptive level, the NP2 target was more likely to attract participants' visual attention when it was uttered by a female (vs. male) speaker.

Figure 17.5

Log-Transformed Fixation Proportions as a Function of NP2 Gender and Speaker Voice by Participants (F_1)



17.4.2 Experiment 4 – The Effects of NP2 Gender and Speaker Voice on Log-Ratios by Items (F_2)

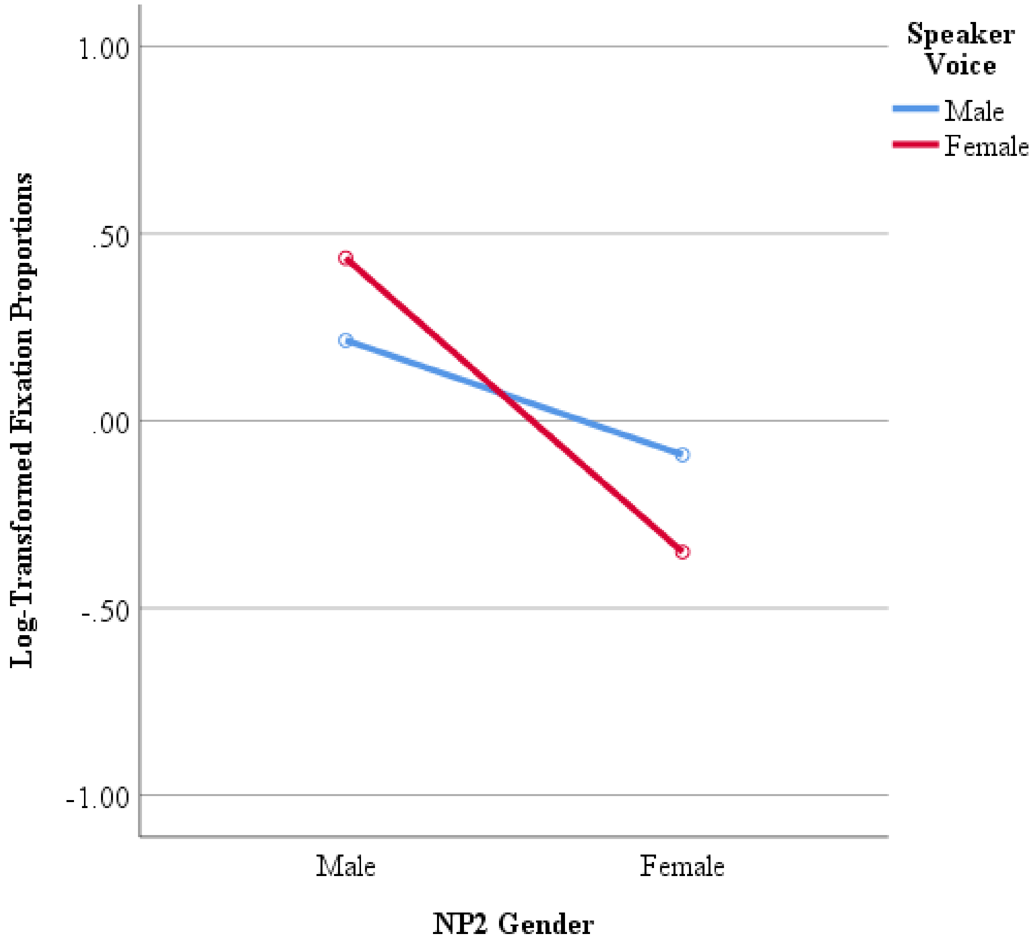
On item-level (F_2), a repeated measures MANOVA was computed with log-ratios calculated by items as a function of NP2 gender as a between-items factor and speaker voice as a within-items factor.

Confirming results found by participants, the main effect of NP2 gender, $F_2(1,56) = 51.41$, $p < .001$, $\eta_p^2 = .479$, was statistically significant, while the main effect of speaker voice was not statistically significant, sphericity assumed, $F_2(1,56) = 0.05$, $p = .828$, $\eta_p^2 = .001$. That is, the NP2 target was looked at, irrespective of the speaker voice. However, contrary to findings by participants, the interaction between NP2 gender and speaker voice was statistically significant in analyses conducted by items, sphericity assumed, $F_2(1,56) = 6.82$, $p = .012$, $\eta_p^2 = .109$.

Figure 17.6 visualizes log-ratios calculated by items as a function of NP2 gender and speaker voice: The fixation pattern was similar to that found in by-participants analyses. As indicated by the statistically significant main effect of NP2 gender, the NP2 target was looked at as soon as it was mentioned. This was independent of whether it was named by a male or a female speaker. Reflecting the interaction between NP2 gender and speaker voice, the NP2 target was more likely fixated when it was named by a female (vs. male) speaker.

Figure 17.6

Log-Transformed Fixation Proportions as a Function of NP2 Gender and Speaker Voice by Items (F_2)



17.5 Experiment 4 – Exploratory Analyses

Exploratory analyses were conducted to investigate the present findings in further detail. Moreover, analogous to Experiment 3, participants' ability to recognize the robots' alleged gender and profession, their ability to imagine robots to perform the presented tasks and professions, and their preferences for robot use were explored. Exploratory analyses can be found in Appendix B, pp. 347.

18 Experiment 4 (Main Verbs - Robots) – Discussion

Experiment 4 was an extended replication of Experiment 2. The main aim was to investigate whether the same results would be obtained when gender-stereotypical main verbs referred to robots. The hypotheses were thus analogous to Experiment 2. Analogous to Experiment 3, the sentence's target was represented by gendered robots.

Contrary to Hypothesis 1, that the main verbs' gender-stereotypicality would guide participants' visual attention, the main effect of main verb gender-stereotypicality was not statistically significant. That is, participants did not look at the robot whose gender matched the main verbs' gender-stereotypicality. Though, at a descriptive level, there was a tendency to look at the gender-matching robot when the main verbs were uttered by a male speaker. This might support the assumption based on Experiment 2 that participants had difficulty to infer the sentence's target from the main verbs' gender-stereotypicality. In Experiment 2, I concluded this difficulty may have arisen because the lines between stereotypically male and female tasks and roles got blurred with changing gender roles (see Auster & Ohm, 2000; Eagly et al., 2019). Testing Hypothesis 2, the main effect of speaker voice was statistically significant only in analyses conducted by items. Contrary to Hypothesis 2, participants did not look at the gender-matching robot when the main verbs were uttered by a speaker whose gender matched the main verbs' gender-stereotypicality. Participants more frequently looked at the robot whose gender matched (vs. mismatched) the speaker's gender instead. A similar tendency was found in Experiment 3 in which adverbs referred to robots. I assumed that participants had difficulty to comprehend that the sentences referred to robots. Participants thus may have used the speaker voice as an indication of the sentence's target. Remarkably, at a descriptive level, when the main verbs were uttered by a male speaker, the preference for the male robot was clearly apparent. When the main verbs were uttered by a female speaker, there was only a weak tendency to look at the female robot. This implies that participants may have been uncertain which robot to fixate when listening to a female speaker voice. Due to their knowledge about robots commonly being male (Jung et al., 2016), they possibly did not expect a female robot to be the sentence's target.

Overall, participants' fixation pattern suggests that they had difficulty to comprehend that the main verbs referred to male and female gender and that the sentence's target was a robot. As such, they could seemingly not infer the sentence's target from either verbal or visual information. It may therefore have required more time and cognitive effort to become familiar with the sentence structure and the visual scenes than in Experiment 1 to Experiment

3. To nevertheless predict the target robot's gender, participants likely used the speaker's gender and their knowledge about robots being male. It is possible that participants who had listened to a male speaker were more sensitive for the main verbs' gender-stereotypicality than those who had listened to a female speaker (see Baron et al., 1991; Barreto & Ellemers, 2005b). At least the tendency to look at the robot whose gender matched the main verbs' gender-stereotypicality when listening to a male speaker might suggest so. However, because participants were seemingly occupied with gaining language comprehension, it is questionable whether they were fully aware of the possibility to face gender stereotypical content. It is therefore likely that they did not fear to possibly behave gender-stereotypical or sexist when listening to the sentences (see Devine, 1989; Monteith, 1993; Monteith et al., 2002). This might explain why contrary to Hypothesis 3a, moreover participants' endorsement of benevolent and hostile sexism and normative gender role orientation did not guide their eye gazes to the robot whose gender matched the main verbs' gender-stereotypicality. Likewise, contrary to Hypothesis 3b, participants' internal and external motivation to control for sexist responses and social desirability did not lead to fewer looks at the gender-matching robot. This was found although participants scored similar on these constructs as participants in Experiment 1 to Experiment 3: Participants scored relatively low on benevolent and hostile sexism, normative gender role orientation, and external motivation to control for sexist responses and relatively high on internal motivation to control for sexist responses and social desirability. This suggests that participants would likely have been motivated to control for sexist responses, if they would have been aware of gender stereotypes and feared to behave gender-stereotypical or sexist. Additional analyses seem to support this assumption.

More concretely, completing the questionnaire, the main effect of participants' endorsement of external motivation to control for sexist responses on perceived female robot gender was just not statistically significant. Likewise, though not statistically significant, participants' endorsement of internal motivation to control for sexist responses tended to affect their levels of robot acceptance. Statistically significant Pearson correlations showed, the higher participants' external motivation to control for sexist responses, the less they indicated to have perceived one of the robots as female. The higher participants scored on internal motivation to control for sexist responses, the less robot acceptance they reported. These findings suggest that participants may have been aware of gender stereotypes and sexism not before the questionnaire items' foci on robot gender, robot acceptance, and sexist attitudes guided their attention to. To illustrate, items on robot acceptance referred to robots as humanlike entities, such as friends, housemates, or employees. They might imply that robots fulfill specific humanlike roles and tasks, such as traditional gender roles. Likewise, female robot gender is often associated with gender stereotypes (Eyssel & Hegel, 2012; Powers et al.,2005; Tay et al.,2014), and sexism (see Strait et al., 2017). It appears thus plausible that particularly items on these constructs might have made participants' motivation to control for sexist responses accessible. During the eye tracking session, however, participants' gazes seemed to be determined by their attempts to obtain language comprehension. This is furthermore implied by their fixation pattern at the end of the sentence when the NP2 target's gender was explicitly named.

More concretely, a statistically significant main effect of NP2 gender indicated that participants succeeded to identify the sentence's target when it was specified. A statistically non-significant main effect of speaker voice indicated that the sentence's target was looked at independent of whether it was uttered by a male or a female speaker voice. The interaction between NP2 gender and speaker voice was statistically significant in analyses conducted by items. In analyses by participants, it was just not statistically significant: Participants were more likely to look at the sentence's target when it was explicitly named by a female (vs. male) speaker. Given the assumption that a female speaker might have indicated that the target robot could be either male or female, participants might have paid more attention to the role's suffix when listening to a female speaker. As participants' expectations about robots seemed to have played an important role over the course of the entire sentence, additional analyses were performed. Participants' fixations at the end of the sentence were investigated as a function of NP2 gender and speaker voice while participants' indications of perceived male robot gender were considered as a covariate. In fact, the interaction between NP2 gender and perceived male robot gender was statistically significant. Pearson correlations showed, the more participants reported to have perceived one of the robots as male, the more they looked at the NP2 target. However, the main effect of NP2 gender turned out not statistically significant when participants' perceptions of male robot gender were considered. That is, perceived male robot gender rather than the explicit mention of the NP2 target seemed to have directed participants' visual attention to the target robot. This contrasts Knoeferle and Crocker (2006; 2007) showing that participants' expectations are placed behind when explicit verbal information is available. As participants' difficulty to establish coherence between verbal and visual information seemed to persist throughout the sentence, they may have preferentially relied on their expectations about robots being male rather than on explicit information to identify the sentence's target.

To better understand participants' expectations, it needs to be considered that most participants reported to know robots from media. This was similar to participants in Experiment

3 (see also Bernotat et al.,2021). As such, the image media create might have coined participants' views of robots (see Bruckenberger et al., 2013; Sandoval et al.,2014). Thus unsurprisingly, participants evaluations of the presented stimuli and of robots in general were similar to Experiment 3. Overall, robots were judged as machinelike and agentic tools. Participants reported moderate levels of robot anxiety and robot acceptance. Their ability to imagine robots to perform the presented tasks and professions was moderate. On the one hand, robotic assistance was appreciated for tasks and professions in which a robot's functions might exceed human capacities, such as managing huge data sets, transporting goods, and working in dangerous settings. On the other hand, participants stated to prefer robots for tasks that do not require close human-robot interaction so that robots could not harm humans (see Bernotat & Eyssel, 2018; Carpenter et al.,2009; Frennert et al., 2013; Horstmann & Krämer, 2019). These evaluations may reflect participants' ambivalence toward robots (Stapels & Eyssel, 2021). Additional analyses uncovered that participants' endorsement of social desirability statistically significantly affected their responses on perceived control over technology. The higher participants' levels of social desirability, the higher levels of control over technology they stated. This indicates that participants may have felt social pressure to report high levels of control over technology. In fact, they apparently perceived less control over technology than they had reported. This might explain their resentments toward robots because perceived control over technology might reduce uncertainty and hesitations to engage in HRI (see also Bernotat et al., 2016).

To sum, in line with Experiment 2, participants seemed to have had difficulty to infer the target's gender from the main verbs' gender-stereotypicality. They furthermore seemed to have had difficulty to comprehend that the sentences referred to robot targets. Their attempts to establish coherence between verbal and visual attention seemed thus to have determined language processing. Participants seemingly used the speaker voice (see Rodriguez et al., 2016; Van Berkum et al., 2008) and their knowledge about robots to predict the target's gender. The latter was apparently relied on even when the target's gender was explicitly specified at the end of the sentence. Only after the eye tracking session, when the questionnaire items may have made participants aware of gender stereotypes and sexism, participants' motivation to control for sexist responses affected their responses.

19 General Discussion

A set of four visual world eye tracking experiments was presented. So far, it seems to be the first theorizing from the fields of psycholinguistics, social psychology, and social robotics. Following different research objectives and approaches, these three fields acknowledged the effects of gender stereotypes on participants' perceptions and behavior. Bridging the fields allowed to gain new insights into the effects of gender stereotypes on language processing and participants' perceptions of robots. To do so, four main research aims were pursued:

First, I investigated whether adverbs' (Experiment 1 and Experiment 3) and main verbs' (Experiment 2 and Experiment 4) gender-stereotypicality would guide participants' visual attention to a target whose gender matched (vs. mismatched) the adverbs'/main verbs' gender-stereotypicality. This seemed promising because the impact particularly of gender-stereotypical adverbs, but also of gender-stereotypical main verbs on language processing was rarely researched. Second, answering Münster and Knoeferle's (2018) call to consider the social context, I investigated whether the effect of gender-stereotypicality would be enhanced by the speaker voice. Third, I have done what is a given in psychology, but still neglected in linguistic research: I took the effects of participants' sexist attitudes, normative gender role orientation, motivation to control for sexist responses, and social desirability on language processing into account. Fourth, I tested whether similar results would be obtained when humans (Experiment 1 and Experiment 2) and when robots (Experiment 3 and Experiment 4) were referred to in language. This was innovative because though robots enter various fields, it was not yet researched how participants' associations with them would affect language processing.

Contrary to the hypotheses, the adverbs' (Experiment 1 and Experiment 3) and the main verbs' (Experiment 2 and Experiment 4) gender-stereotypicality and the speaker voice did not guide participants' visual attention to a target whose gender matched the adverbs' or main verbs' gender-stereotypicality. Nevertheless, gender-stereotypicality and speaker voice seemed to have affected language processing to some extent. More specifically, in Experiment 1, participants looked at the character whose gender matched the adverbs' gender-stereotypicality when listening to a female speaker voice. In Experiment 3, the hypothesized fixation pattern to look at the gender-matching character, particularly when the speaker's gender matched the adverbs' gender-stereotypicality, was found in tendency. Taken together, the results of all four experiments imply that language processing was seemingly driven by two motives: First, participants apparently attempted to gain language comprehension. Going beyond prior linguistic literature, it seems that, after participants comprehended that the sentences referred to a male or a female target, they apparently attempted to respond in a non-stereotypical and non-sexist manner. This was indicated by counteractive fixations on the stereotype-inconsistent character. In addition, the results of Experiment 3 and Experiment 4 gave insights into participants' views of robots. The results indicate that participants' stereotypes about robots per se as well as their attitudes toward the group robots represented in terms of gender seemed to have affected participants' cognitions during language processing.

19.1 Gaining Language Comprehension

The results of Experiment 1 and Experiment 3 suggest that it did not require much time and cognitive effort to comprehend that the adverbs referred to either of the displayed characters. In Experiment 1, this was indicated by looks at the character whose gender matched the adverbs' gender-stereotypicality when the adverbs were uttered by a female speaker. In Experiment 3, this was indicated by a tendency to look at the robot whose gender matched the adverbs' gender-stereotypicality. The adverbs had presumably automatically activated stereotypical associations with men and women. In comparison, in Experiment 2 and Experiment 4, there was no clear tendency to look at the target whose gender matched (vs. mismatched) the main verbs' gender-stereotypicality. Fixation patterns thus indicated that participants might have had difficulty to infer the target's gender from the main verbs' gender-stereotypicality. This led to the assumption that the adverbs may be stronger associated with gender than the main verbs. This appears plausible because people infer men's and women's attributes from their actions and roles (Eagly et al., 2019; Koenig & Eagly, 2014; Wood & Eagly, 2012). Men's and women's roles have changed drastically (see Auster & Ohm, 2000; Eagly et al., 2019). Associations between attributes and gender may thus be stronger because they persist for a longer time and are more frequently activated than associations between actions and gender (see Fazio, 1990; 2007; Fazio et al.,1982 for association activation). Nonetheless, to easily link verbal and visual information both apparently needed to match participants' expectations (see also Huettig & Altman, 2005). Because adverbs and main verbs commonly describe humans, participants in Experiment 3 and Experiment 4 did apparently not expect robots to be the sentence's target. Therefore, particularly in Experiment 4, participants seemed to have had difficulties to predict the sentence's target because neither the main verbs nor the visual scenes showing robots allowed to infer its gender. To nevertheless predict the sentence's target, participants in Experiment 3 and Experiment 4 seemed to have used the speaker voice (see also Van Berkum et al., 2008) and their knowledge about robots commonly being male (Jung et al., 2016). This was indicated by a general preference for the male robot, particularly when the speaker voice was male.

Taken together, the results can be interpreted in the context of the Social CIA (Münster & Knoeferle, 2018; see also Knoeferle & Crocker, 2006; 2007; Knoeferle et al., 2014): The interplay of verbal and visual information, the social context in terms of the speaker voice, and prior knowledge seemed to have determined language comprehension. Early verbal information, such as the adverbs (Experiment 1 and Experiment 3) and the main verbs (Experiment 2 and Experiment 4), seemed to be used to predict future incoming verbal information (see also Altman & Kamide, 1999; Kamide et al.,2003). Thus unsurprisingly, it seems that, the less effort it took to infer a target's gender from the adverbs' or the main verbs' gender-stereotypicality, the easier it seemed to identify the NP2 target when it was explicitly named at the end of the sentence. In case the NP2 target's actual gender mismatched participants' expectations due to the anteceding adverbs' or main verbs' gender-stereotypicality, participants apparently revised their initial prediction about the NP2 target's gender. This was reflected by corrective eye movements from the target they anticipated to the one that was named. This fixation pattern confirms that language processing occurs incrementally word by word (see also e.g., Altman & Kamide, 1999; Kamide et al., 2003; Pyykkönen et al., 2010; Rodriguez et al., 2016). Furthermore, the findings demonstrate that, if possible, participants preferentially rely on explicit verbal information rather than on their expectations (see Knoeferle & Crocker, 2006; 2007). This seems logical because usually explicit verbal information reliably specifies a target's gender. A validation of this information by referring to gender stereotypes and prior knowledge is thus not necessary for a correct target identification.

19.2 Attempts to Respond Non-Stereotypical and Non-Sexist

The ease with which the target's gender could be inferred from the adverbs', respectively, the main verbs' gender-stereotypicality that seemed to have facilitated language comprehension seemingly furthermore determined whether and how promptly participants considered likely to encounter gender-stereotypical content. This possibly occurred after some trials because the sentences' structure, the male (vs. female) speaker voice, and the visual scenes displaying a male and a female target remained constant during the experiment. It may have led participants to reflect their fixation behavior and to reconcile it with their attitudes toward sexism and their willingness to behave non-sexist and socially appropriate (see Monteith, 1993; Monteith et al., 2002; see also Gawronski & Bodenhausen, 2006; 2011). More precisely, on the self-report measures, participants indicated relatively low levels of sexism and relatively high levels of internal motivation to control for sexist responses and social desirability across all four experiments. Accordingly, the encounter of gender stereotypes and the sense to possibly have behaved gender-stereotypical or sexist during past experimental trials likely evoked feelings of discomfort which participants might have attempted to resolve. To do so, their vigilance for indications of gender stereotypes and sexism might have been enhanced in following trials (see Monteith, 1993; Monteith et al., 2002). To identify indications of gender stereotypes, participants might have consciously paid attention to the experimental setting and conjectured about the purpose of the experiment. Their attempts to counteract gender stereotypes and sexism seemingly resulted in less fixations on the stereotype-matching target. Fixations on the gender-matching and the stereotype-inconsistent target were interrelated (see Arai et al., 2007). Therefore, avoiding to look at the gender-matching target led to fixations on the stereotype-mismatching one. Student samples were found to indicate relatively low levels of sexism (De Judicibus & McCabe, 2001; Pettijohn & Walzer, 2008). This led to the assumption that responding non-sexist may be regarded a norm amongst students. If this assumption was correct, participants – who were students – may have internalized the norm to respond non-sexist. They may have 'learned' to counteract gender stereotypes and sexism whenever they encounter them. As such, participants could have automatically avoided to look at the character that matched the adverbs' or main verbs' gender-stereotypicality when language comprehension was obtained effortlessly (see De Houwer, 2009; Devine, 1989). Therefore, the effects of participant gender amongst participants' sexist attitudes and the interaction between the adverbs' gender-stereotypicality and speaker voice were particularly apparent in Experiment 1. This seemed plausible because in Experiment 1, language comprehension seemed to be gained easiest because the adverbs that were supposed to be strongly associated with gender referred to male or female human targets. Therefore, participants in Experiment 1 might have paid more attention to contextual factors, such as their own gender, the speaker voice, and, in tendency, the adverbs' connotation than participants in the remaining experiments. Consequently, they apparently countered the adverbs' gender-stereotypicality by avoiding fixations on the character whose gender matched the adverbs' gender-stereotypicality, particularly when the speaker and participants themselves were male. Likewise, in Experiment 1, participants indicated lower levels of benevolent and hostile sexism and normative gender role orientation after having listened to a speaker of their opposite gender. These findings led to the assumption that participants' and the speaker's group membership in terms of gender played a crucial role when being exposed to gender-stereotypical language that refers to male and female targets. These findings furthermore suggest that social, particularly non-sexist norms might have been activated during the experiment. Strengthening this assumption, participants' external motivation to control for sexist responses led to more looks at the robot whose gender mismatched the adverbs' gender-stereotypicality in Experiment 3. As such, the context of the experiment taking place on a university campus with the experimenter, a student peer, being around had possibly activated non-sexist norms that might prevail amongst students.

Remarkably, in Experiment 1, participants' attempts to counter gender stereotypes when listening to the adverbs seemed to have caused a "bend over backward" effect (Eagly & Mladinic, 1989, p. 554) – The reflection of particularly positive associations with women. This was indicated by a tendency to preferably look at the female character rather than the male one, especially when positively connoted adverbs were uttered by a male speaker. Though the interaction between adverb connotation and speaker voice was only statistically significant in by-participants analyses. In comparison, in Experiment 2, in which participants seemed to have difficulty to infer the target's gender from the main verbs' gender-stereotypicality, participants tended to look at the character whose gender matched the main verbs' gender-stereotypicality when listening to a speaker of their own gender. They tended to look at the stereotype-inconsistent character when the main verbs were uttered by a speaker of their opposite gender. Particularly in Experiment 1 and, to a lesser extent, in Experiment 2, the speaker voice and participant gender may thus have served as an indication of gender stereotypes. Recall that in Experiment 3 and Experiment 4, in which participants apparently needed to comprehend that the sentence's target was a robot, the speaker voice seemed to be used to infer the target's gender. This suggests that participants' motive during language processing – gaining language comprehension or responding non-stereotypical and non-sexist – may have determined how promptly information, such as the speaker's gender, was integrated and how it affected their eye gazes.

19.3 Participants' Views of Robots

Apart from participants' language processing, the results of Experiment 3 and Experiment 4 suggest that socially shared views and expectations seemed to have also played a role in the context of robots. More precisely, according to responses on the self-report measures in both experiments, the majority of participants knew robots from media. It is thus likely that media have coined participants' views of robots (see Horstmann & Krämer, 2019; Sandoval et al.,2014). At least participants' evaluations of robots as machine-like, agentic tools that were rather associated with male gender and thus preferred for stereotypically male tasks might suggest so (see also Bernotat & Eyssel, 2018; Bernotat et al.,2017; 2021). Stereotypically male tasks were judged as requiring less close interaction with a robot (Bernotat et al.,2021). Accordingly, robots were preferred for assistive tasks in which their capacities may exceed humans' or which may be dangerous for humans, but which do not require close HRI (see also Bernotat & Eyssel, 2018; Frennert et al.,2013; Horstmann & Krämer, 2019). These preferences possibly reflect trust in a robot's functions, amongst doubts in its benevolence (cognitive vs.

affective trust, Bernotat et al.,2017; 2021). Perhaps participants were ambivalent about robots which is also suggested by their moderate levels of self-reported robot acceptance and robot anxiety (Stapels & Eyssel, 2021).

Due to participants' views of robots as machine-like tools, it seems logical that their ability to imagine robots to perform the presented tasks and professions, that are commonly performed by humans, was limited. Likewise, attributions of machinelikeness, agency, and male robot gender seemed to be inconsistent with the presentation of female robots that were addressees of gender stereotypical content. Therefore, a female robot being displayed amongst a male one during the eye tracking session may have directed participants' attention to gender stereotypes and sexism. As such, participants likely needed to match their perceptions of and attitudes toward robots with those they shared toward gender stereotypes and sexism. This may support the assumption that robots are more than representatives of humanlike categories, such as gender (Bernotat et al.,2017; 2021; Eyssel & Hegel, 2012; Otterbacher & Talias, 2017) and ethnicity (Eyssel & Kuchenbrandt, 2012; Eyssel & Loughnan, 2013; Makatchev et al.,2013). Robots seemed to be perceived as a distinct group or category (see also Bernotat et al.,2021; Wullenkord & Eyssel, 2020b). Therefore, social expectations seemed indeed to play a role in the context of robots.

According to participants' visual attention and their questionnaire responses, they strived to behave non-sexist toward any entity, including robots. At the same time, that social desirability led to lower levels of robot acceptance in Experiment 3 implies that participants seemingly deemed socially appropriate not to accept robots. This may have been due to the image of robots that media create. Perhaps participants regarded technological development in general with skepticism. This is suggested by the fact that participants' levels of technology acceptance were moderate to low, though they indicated a good command of technology use.

19.4 Considerations for the Interpretation of the Present Results

The present results may certainly enrich research on language processing of gender stereotypes and social robotics. Nevertheless, to understand the present findings correctly, it needs also to be acknowledged that results provided by implicit measures, such as eye tracking, should always be interpreted with caution. More specifically, participants' cognitions during language processing were inferred from their gaze behavior when watching the visual scenes and listening to the sentences. As common for implicit measures, participants' gazes may have been affected by incidental features of the visual and the verbal stimuli that are hard to detect and thus to control (see Bluemke & Friese, 2006; see Fiedler et al.,2006; Gawronski & Hahn, 2019; Unkelbach & Fiedler, 2020 for reviews). To illustrate, some of the characters (taken from

Guerra et al., 2021) that were displayed in Experiment 1 and Experiment 2 were frontally displayed, others were laterally shown. Some characters had a dark skin color while others had a light one. Therefore, the characters may have attracted participants' visual attention differently due to differences in their visual features. Due to differences in skin color, the characters could also have activated associations with and attitudes toward ethnic groups which may have affected participants' responses. These are just examples. To avoid such differences in the visual features, the robot stimuli used in Experiment 3 and Experiment 4 were kept constant apart from differences in WHR and SW (Bernotat et al.,2017; 2021). Keeping stimuli as constant as possible in size and style across experiments, pretesting them, using pseudo-randomized item lists, and analyzing data by participants and by items may have helped to diminish incidental stimuli effects (see Arai et al., 2007; Clark, 1973; Pollatsek & Well, 1995). In addition, data analyses by participants and by items allowed to treat each experimental factor (gender-stereotypicality, speaker voice) once as a within- and once as a between-subjects/items-factor. This ought to diminish an imbalance due to the fact that in mixed designs the within-factor naturally gains more data points and is thus more precisely measured than the between-factor. Nonetheless, that way, conclusions were drawn from different observation levels in terms of by-participants and by-items analyses whose outcomes sometimes differed within-experiments. Considering the covariates, possible explanations for some differences between by-participants and by-items results could be provided. Furthermore, conclusions were drawn from comparisons within- and between-experiments. According to meta-analyses, levels of heterogeneity across experiments were moderate by participants and by items. This might have been due to differences in sample sizes because of drop outs, speakers' reading styles (despite training), the fact that, unlike the adverbs' connotation, the main verbs' connotation was not considered and that, unlike the robots, the human targets also differed in posture and skin color. Moreover, post-hoc analyses of statistical power using sample sizes by participants and by items and SPM estimates for sample sizes required to obtain .80 statistical power in future replication experiments suggest that statistical power was relatively low in all four experiments. That is, even if adverb and main verb gender-stereotypicality and speaker voice might have affected participants' visual attention, it is very likely that their effects did not turn out statistically significant due to low statistical power (see Stevens, 2009). Participants found to look at the gender-matching (vs. mismatching) character dependent of the speaker voice in Experiment 1, participants' tendency to show the hypothesized fixation pattern in Experiment 3, and the tendency to look at either of the robots dependent of the speaker voice in Experiment 3 and Experiment 4 might support this assumption. Furthermore, though not in main focus of investigation, it needs to be considered that IDAQ scale showed low internal consistencies in Experiment 3 and Experiment 4. Thus, as valuable the insights provided by the present research are, the conclusions are tentative. Moreover, as every research, the scope of this line of research was limited. Conclusions drawn from the present results thus need to be confirmed and extended in follow-up research. The best way to do so, may be to tie in with the limitations of the present research.

19.5 Future Work on Language Processing and Gender-Stereotypes

Monteith's *self-regulation model* (see Monteith, 1993; Monteith et al., 2002) was referred to in order to infer participants' cognitions during language processing. Due to the speaker voice, the sentence structure, and the array of the visual scenes remaining constant across the experiment, participants were supposed to have felt discomfort to possibly behave gender-stereotypical or sexist after some trials. Particularly in Experiment 1 in which language comprehension seemed easy, participants were supposed to have thus countered fixations on the character that matched the adverbs' gender-stereotypicality when listening to a male speaker voice to resolve this discomfort. If this assumption was correct, participants' emotional arousal likely increased during the experiment when they feared to behave gender-stereotypical or sexist. This further suggests that participants' fixation pattern may have changed when participants feared to behave sexist and avoided fixations on the gender-matching character accordingly. The present findings however do not reflect participants' arousal during the experiment and changes in fixation patterns were only descriptively displayed by the time course graphs. Therefore, an extended replication of Experiment 1 may be done. Extending Experiment 1, measures of fixation proportions may be complemented by measures of fixation duration, saccade proportion, and pupil diameter. More specifically, Monteith (1993; Monteith et al., 2002) stated that counteractive processes were preceded by a behavioral inhibition. If the findings from Experiment 1 can be replicated and if they can indeed be explained by Monteith's model, such an inhibition may be reflected by participants' eye movements: That moment participants feel discomfort to possibly behave gender-stereotypical, likely their pupil dilatation increases as an indication of emotional arousal (Ren et al.,2014; Zhai & Barreto, 2006). At the same time, their eye gazes possibly reside or repeatedly switch between the two characters before the stereotype-inconsistent character is fixated. This may lead to an increase of pupil dilatation, fixation duration, and saccade proportion also when listening to the adverbs in following trials (see Monteith, 1993; Monteith et al., 2002). Before this increase, likely the character whose gender matches the adverbs' gender-stereotypicality is fixated, independent of the speaker voice, because its gender is automatically associated with the adverbs' stereotypicality. After this increase, likely the stereotype-inconsistent character is fixated. Likewise, after this increase, participants' sexist attitudes, motivation to control for sexist responses, and social desirability may more strongly affect their gazes than before. In the present experiments, the interactions between participants' external motivation to control for sexist responses with adverb gender-stereotypicality was statistically significant only in Experiment 3. Although, participants' fixation patterns suggest that participants attempted to counter gender-stereotypical language and sexism. Perhaps the effects of participants' internal and external motivation to control for sexist responses and social desirability may have been more apparent in the present experiments if it would have been possible to examine fixation patterns before and after this increase. Therefore, investigating fixation patterns before and after this increase may provide more detailed information about the impact of participants' sexist attitudes, motivation to control for sexist responses, and social desirability on their eye gazes during language processing of gender-stereotypical content.

Measuring pupil dilation and fixation duration could also be useful to test the assumption that language processing was less cognitively demanding when gender-stereotypical adverbs (vs. main verbs) referred to human (vs. robot) targets. More concretely, in an early research, Just and Carpenter (1993) found that fixation duration and pupil dilatation increased with sentence complexity and thus reflected cognitive load during language processing. Pupil dilatation has also shown to be a useful measure of cognitive load during an HRI task (see Ahmad et al.,2019; Minadakis & Lohan, 2018). Conducting extended replications of Experiment 1 to Experiment 4 while measuring cognitive load and fixation duration may confirm whether gaining language comprehension was more cognitively demanding when main verbs (vs. adverbs) and robots (vs. humans) were presented. Moreover, it could be tested whether cognitive load distracted participants from countering fixations on the target whose gender matched the adverbs' or main verbs' gender-stereotypicality. However, it needs to be considered that pupil dilatation also increases under emotional stress (Ren et al., 2014; Zhai & Barreto, 2006). The exposure to gender-stereotypical language likely evokes emotional stress, irrespective of participants' difficulty to comprehend the sentences. Therefore, participants' level of perceived psychological stress should be considered, e.g., by using Lepore et al.'s (1993) scale.

Regarding participants' attempts to gain language comprehension, the speaker voice was assumed to be integrated differently when it was relevant for language comprehension as when it served to counter gender stereotypes. To confirm this assumption, the speaker voice's relevance for target prediction may be examined in future research. To illustrate, Van Berkum and his colleagues (2008) presented sentences spoken in the 1st person form singular (e.g., "If only I looked like Britney Spears in her latest video." uttered by a man vs. woman). The 1st person form singular "I" implied that the speaker was likely to be the sentence's target.

The speaker's characteristics (e.g., gender) were thus relevant to anticipate the sentence's target. Using the 3rd person form singular in passive voice, no such direct link between the speaker and the sentence's target existed in the present experiments. That way, the speaker voice could have been less relevant for gaining language comprehension than expected. This could explain why the speaker voice did not guide participants' visual attention to the target whose gender matched the adverbs or main verbs gender-stereotypicality as hypothesized, while in prior research it did (see Van Berkum et al., 2008; see also Rodriguez et al., 2016 for similar findings using a target's hands). To test this assumption, the gender-stereotypical adverbs and professions from Experiment 1 may be embedded in 1st (vs. 3rd) person form sentences that vary between-subjects, e.g., "*I carefully* watched the window because I am a *construction worker*." (1st form) vs. "Someone *carefully* watched the window because *he* (vs. *she*) is a *construction worker.*" (3rd form). Such sentences would moreover bear the advantage of having a canonical sentence structure, a structure that is common in every-day language. Furthermore, varying the speaker voice as a within-subjects factor, the sentences may be read by a male (vs. female) speaker. At the same time, the visual scenes used in Experiment 1 may be displayed. In addition, after the experiment, participants may be asked whether they considered likely that the speaker was the sentence's target and to complete items on ambivalent sexism and motivation to control for sexist responses and social desirability. Those participants who listen to the 1st form sentences may more likely consider the speaker to be the sentence's target than those who listen to 3rd form sentences. If so, participants who listen to 1st (vs. 3rd) form sentences may consider the speaker' s gender relevant to predict the sentence's target similar to participants in Van Berkum and colleagues' (2008) research. Therefore, participants who listen to 1st (vs. 3rd) form sentences may look at the character whose gender matched (vs. mismatched) the adverbs' gender-stereotypicality irrespective of whether the sentences were uttered by a male or a female speaker. In comparison, because participants who listen to the 3rd (vs. 1st) form sentences presumably consider the speaker less likely to be the sentence's target, particularly a male speaker voice may draw their attention to gender stereotypes (see Baron et al., 1991; Barreto & Ellemers, 2005b). That is, similar to participants in Experiment 1, participants who listen to 3rd form sentences may pay attention to the speaker voice that alters between trials and to the adverbs' gender-stereotypicality and the visual scenes. They may be more likely to suspect the experiment being related to gender-stereotypes than participants who listen to 1st form sentences. Thus, they may look at the target whose gender mismatches (vs. matches) the adverbs' gender-stereotypicality, particularly when listening to a male speaker voice similar to participants in Experiment 1 did. Likewise, the effects of participants' sexist attitudes, motivation to control for sexist responses, and social desirability on their eye gazes when listening to the adverbs may be stronger for participants listening to 3rd (vs. 1st) form sentences.

Apart from concrete future experiments, some recommendations can be derived from the present experiments that may be valuable for future visual world experiments on language processing of gender stereotypes. To illustrate, in the present experiments, the visual stimuli were kept as constant as possible to avoid unintended variation. However, a male (vs. female) speaker voice, the sentences ending on a male (vs. female) professional role, and a male amongst a female character being displayed across trials seemingly directed participants' attention to gender stereotypes and sexism. Therefore, in follow-up research, the filler sentences should differ from the experimental sentences. Not to point to gender stereotypes and sexism, there should be a notable, but not too obvious difference between the experimental and the filler sentences. Therefore, the sentence structure in terms of active or passive voice should be the same for filler and experimental sentences. However, to conceal the experiment's relation to gender stereotypes, the filler sentences should not contain either a male or a female target, if the experimental sentence's target would be male or female as in the present experiments. To illustrate, "All planets are embraced by the universe." may be a suitable filler sentence because Bernotat, et al. (in preparation) found the universe being evaluated as neutral with regard to gender stereotypes. The visual scenes related to the fillers would naturally differ from the experimental trials because no male and female characters would need to be depicted. Moreover, if the speaker voice would vary within- (vs. between-) subjects and if the sentences would be read by various male and female speakers, participants would possibly pay less attention to the speaker voice. The generalizability of the effect of the speaker voice on language processing would also be increased. In case researchers were interested in what participants were consciously paying attention to, participants could be asked to speak their thoughts and impressions aloud during the experiment (see Davison et al.,1995 for a suitable think-aloud task).

With regard to Münster and Knoeferle's (2018), the present results demonstrate that the social context needs to be taken into broader perspective in future research on language processing. This is indicated by the context of the experiments taking place on a university campus with the experimenter, a student peer, being next to the participant that was supposed to have made sexist attitudes and non-sexist norms accessible. It needs to be considered that social norms are interrelated with demographics of all who are involved in language processing, such as participants', experimenter's, and speaker's group membership in terms of gender, professional status, age, and ethnicity.

In this regard, participants in the present experiments indicated relatively low levels of sexism as common for students (see De Judicibus & McCabe, 2001; Pettijohn & Walzer, 2008). To draw conclusions about the impact of gender-stereotypical language and sexist attitudes on visual attention during language processing with more confidence, samples should be tested whose sexist attitudes and motivation to control for sexist responses range more widely. Therefore, samples should differ in age, ethnicity, and socio-economic background with sample sizes by participants and by items being sufficiently large. Though, realizing large sample sizes by-participants has shown to be difficult in eye tracking research because many people are just not willing to participate in an eye tracking experiment in which they have to keep their head in a chin rest. Moreover, dropout rate can be high due to technical or calibration issues. Likewise, increasing sample sizes by-items is also challenging due to complex and time-consuming pretests needed to gain large samples of suitable items.

Moreover, future research needs to consider that using the visual world paradigm, it remains unclear which effects on participants' cognitions (i.e., shifts of visual attention, activation of attitudes and thoughts) were actually caused by the adverbs or main verbs. Therefore, eye tracking data needs to be complemented by other experimental approaches.

To illustrate, the fact that the gender-stereotypical adverbs and main verbs were embedded in a whole sentence in the present experiments was advantageous because this resembled natural language use. However, the noun anteceding the adverb and the main verb, respectively, could have been followed by any verb related to the noun, while an adverb preceding the verb is uncommon in natural language use. At the same time, the visual scene portrayed neither the main verb nor the adverb. Thus, to predict the target of the sentence, participants had to listen until the adverb or main verb was fully articulated which took about two seconds each (plus the time before its onset). Considering that automatic cognitive processes, such as the activation of attitudes and motives and shifts of visual attention occur within only a few milliseconds (see Allopenna et al., 1998; Fazio, 2007; Tanenhaus et al., 1995; see also Schneider & Chein, 2003 for a review), this time window of about two seconds is relatively long. It is thus unclear whether sexist attitudes and motives were activated by the adverbs or main verbs or even already before their onset. Moreover, given this relatively large time window and the fact that adverbs or main verbs were embedded in a whole sentence, the conclusion that adverbs may be more strongly associated with gender than main verbs is tentative and needs to be confirmed in follow-up research. The same accounts for the assumption that positive adverbs and main verbs may be more promptly associated with female gender than with male gender because participants attempt to respond non-sexist.

Results from eye tracking data could therefore be complemented and extended by using the *Implicit Association Test* (IAT; Greenwald et al.,1998) or the *Sequential Priming Task* (see Gawronski & De Houwer, 2014; Gawronski & Hahn, 2019 for reviews). Both tasks would enable to present adverbs or main verbs in isolation. Using different intertrial intervals or *Stimulus Onset Asynchronies* (*SOA*), participants' implicit responses may be investigated in times ranging from subliminal priming to ones that allow for conscious awareness and control (see Castillo-Mayén & Montes-Berges, 2017; Greenwald et al.,1995; Koch & Tsuchiya, 2007 for possible timings).

An IAT may be conducted as follows: Blocks may contain the stereotypically male (vs. female) adverbs and main verbs from the present research and words with a positive (vs. negative) connotation. Negative and positive words may be taken from the *Berlin Affective Word List* (BAWL-R, Võ et al.,2009)[38]. Moreover, according to Greenwald and colleagues (1998), the time between participants' response and the following trial may be varied systematically ranging from about 100ms to 700ms. Unlike the present experiments, amongst the adverbs' connotation, the main verbs' connotation should be balanced because positivity and negativity of stimuli have shown to play a role in prior IAT experiments (e.g., Bluemke & Friese, 2006; Govan & Williams, 2004).

Regarding the sequential priming task, a scenario according to Castillo-Mayén and Montes-Berges (2017) would allow to mirror participants' perceptions of each the adverbs' and the main verbs' gender-stereotypicality and participants' perceptions of the adverbs and the main verbs' connotation. To illustrate, either simply the written words "male" vs. "female" or, alternatively, the images of the male and the female characters that were used in the present research may be presented as prime stimuli. The prime stimulus may be followed by a mask which would then be followed by the adverbs and the main verbs from the present research used as target stimuli. Participants would then have to indicate whether the target was positively or negatively connoted. Shorter response times for congruent (vs. incongruent) trials (congruent: e.g., the word "male" or image of a male character followed by a stereotypically male adverb or main verb, incongruent: e.g., the word "male" or image of a male character followed by a stereotypically female adverb or main verb) would indicate whether the adverbs and the main verbs were associated with male or female gender. Participants' responses in terms of

[38] The BAWL-R (Võ et al.,2009) contains a large set of German words, their psycholinguistic indexes known to influence word processing, and ratings on the words' emotional arousal, valence, and imageability. This allows to control for a wide range of factors that may affect language processing apart from the experimental manipulation.

connotation may indicate whether the adverbs and the main verbs were regarded as positive or negative.

The effects of participants' sexist attitudes and motivation to control for sexist responses on their response times need to be considered as covariates similar to the present experiments. To document when and to what extent participants' sexist attitudes and motivation to control for sexist responses affected their responses over time, the covariates' impact should be measured per intertrial interval or SOA, respectively. To test the assumption that the adverbs are more strongly associated with gender than the main verbs, the results of the proposed IAT and the sequential priming task may be analyzed for adverbs and main verbs separately. Participants' evaluations of the adverbs' and main verbs' connotation may confirm whether female gender is more strongly associated with positive words than male gender and whether this reflection of particularly positive associations with women was affected by participants' motivation to control for sexist responses.

19.6 Future Work on Social Robotics

The results of Experiment 3 and Experiment 4 suggest that insights into participants' perceptions of robots and, in this regard, particularly reasons for participants' hesitations toward robots need to be deepened in future research. More specifically, in the present experiments, participants' preferences to use robots suggest that participants would hesitate to interact with robots. Their hesitations presumably resulted from an ambivalent view of robots that was likely coined by the picture media portrayed of robots as being supportive on the one hand and being threatening on the other (see Horstmann & Krämer, 2019; Sandoval et al., 2014). This led to the assumption that participants had high trust in a robot's functions (cognitive trust), but low trust in its benevolent motives (affective trust, Bernotat et al., 2017; 2021). In addition, participants apparently considered socially appropriate not to accept robots, which may have reinforced their reluctance toward robots. The present findings suggest a relation between hesitations, ambivalent views of robots, and cognitive and affective trust toward robots. Nonetheless, answering whether such a relation exists would have exceeded the scope of the present research and is thus up to follow-up experiments. To do so, an *Approach-Avoidance-Task* (Chen & Bargh, 1999; see Gawronski & Hahn, 2019 for a review) may be done: In one block, participants may be instructed either to approach or to avoid a given stimulus, e.g., a picture of a robot (vs. human vs. neutral stimulus; see Stapels & Eyssel, 2021 for possible neutral stimuli). In another block, words related to a robot's functions (vs. motives) may be presented (these may be adapted from items on cognitive and affective trust, Bernotat et al.,2017; 2021). Instead of instructing participants to move a lever, movements of the mouse cursor to a given stimulus (vs. to a distant point) may be tracked (see Stillman et al.,2018). Reflecting participants' hesitations toward robots, participants may be faster and use a straighter trajectory when getting far from (vs. close to) a robot (vs. humans vs. neutral) stimulus and when words were related to a robot's motives (vs. functions). At the same time, participants' ambivalence toward robots (Stapels & Eyssel, 2021), cognitive and affective trust (Bernotat et al.,2017; 2021), and their social desirability may be assessed. That way, it may be confirmed whether ambivalent feelings toward robots and trust in robots' functions amongst a low confidence in their benevolence prevent participants from approaching robot stimuli. Considering the effects of social desirability may uncover whether participants' hesitations to approach robot stimuli were driven by attempts to behave socially appropriate.

In the present research, 'typical' humanlike robots were deliberately used because it seemed to be the first research on language processing featuring robots. For pioneering research, the use of typical robots was considered advantageous because this way participants likely recognized the robots as such and had the same image of robots in mind during language processing. However, future research should consider that robots increasingly enter the market that do not appear like what participants in Experiment 3 and Experiment 4 considered a typical robot. For instance, hoover and mowing robots do not come along with arms, legs, a torso, and facial cues. Because these commercial non-humanlike robots already entered people's daily lives, they should be set in focus of future investigations. Future research should examine to what extent commercial non-humanlike robots were considered belonging to the category of robots and whether social norms play a role for robot perception. To do so, participants could be presented with a set of typical humanlike (vs. commercial non-humanlike) robots which may be presented one after another on a screen. Typical humanlike robots can be used from Experiment 3 and Experiment 4. At the same time, participants' pupil dilatation may be measured to reflect their emotional arousal (see Ren et al., 2014). Participants may indicate to what extent they count the displayed robot to the category of robots. This may be done by adapting the *Inclusion of Other in the Self Scale* (Aron et al.,1992). Moreover, participants may judge the robot's gender. In addition, they would complete items on social desirability and external and internal motivation to control for sexist responses.

Perhaps participants' perceptions of robots as male do not apply to commercial non-humanlike robots. That way, an effect of participants' external motivation to control for sexist responses may be less apparent when commercial non-humanlike (vs. typical humanlike) robots were presented. At the same time, commercial non-humanlike robots are likely less strongly associated with the image of robots as being threatening because participants may have experienced them as domestic tools in their daily lives. Therefore, participants may experience less cognitive arousal toward commercial non-humanlike (vs. typical humanlike) robots. Moreover, they may thus not consider socially expected not to accept commercial robots.

Research on commercial robots is needed because researchers still seem to hold on to the idea of humans living with robot companions, while developers and investors apparently understood what the present data on Experiment 3 and Experiment 4 suggest: People want robotic assistance as long as robots come along as useful tools they can easily control (see also Bernotat & Eyssel, 2018; Schiffhauer et al., 2016). Instead of being guided by stereotypes about robots themselves, researchers should adapt to the idea that, if at all, humans will very likely be surrounded by robots that appear as technical devices or mediums that serve and connect people instead of replacing them. Promoting and realizing this view of robots should be roboticists main aim as it might be the key for a useful and constructive HRI.

19.7 Contributions of the Current PhD Project

This interdisciplinary research project combined theories from social psychology, psycholinguistics, and social robotics and thus contributed to new insights and perspectives in all of these fields. Emphasizing that participants' sexist attitudes and motives, such as their attempts to counter gender stereotypes and sexism, might crucially determine language processing represents a major contribution. So far, the impact of the social context and of participants' beliefs and attitudes on language processing was acknowledged in prior psycholinguistic research (e.g., Altman & Kamide, 1999; Münster & Knoeferle, 2018; Van Berkum et al., 2008; 2009). However, this particular line of research did not yet consider that participants' attitudes and motives may affect their visual attention in its own right after language comprehension was obtained. As such, the present results suggest that participants' eye movements during language processing may not necessarily merely reflect their attempts to understand and to predict incoming language. This notion does not question the eye tracking method's validity, it rather calls to broaden the scope of future psycholinguistic research by bearing in mind that participants' eye movement data may reflect their attempts to be consistent with their attitudes, motives, and social norms during language processing. These seem to be closely related to the social background of all who are involved in language processing, such as listeners' or readers' and speakers' group membership in terms of gender, ethnicity, age, and socio-economic status (see also Rodriguez et al., 2016; Van Berkum et al., 2008). Researchers need to consider that participants' attempts to be consistent with their attitudes, motives, and social norms in turn seem to depend on association strength, awareness of stereotype content, cognitive resources, time, and the level to what attitudes, norms, and behavior are internalized and thus precede automatically (see Conrey et al., 2005; Devine, 1989; Fazio, 1990; 2007). This further implies that automatic and controlled processes seem to work in parallel during language processing. All these aspects remained mainly unattended in prior psycholinguistic research. Emphasizing their impact on language processing is a great value of this research project.

Likewise, because gender stereotypes are usually passed on in verbal communication, investigating their effects on language processing likely broadened social psychologists' view of gender stereotypes as a linguistic phenomenon. An additional benefit of the present research lays in the evaluative processes that were needed to create the sentences and the visual scenes. Those go widely beyond the mere preparation of experimental stimuli. A vast set of tasks (> 140 in total), attributes (168 in total), and nouns (> 140 in total) was provided. These items precisely picture what is nowadays considered stereotypical for men and women and which attributes are regarded as positive and negative in large German samples.

In the context of research on social robotics, a crucial research gap was closed by evaluating a set of 14 robots that are currently placed on the market on robot typicality, recognizability, suitability for tasks and professions, and robot gender. This might seem trivial. However, because it is generally assumed that everyone knows robots or at least has an idea about them, empirical data on which real existing robot types are actually considered typical robots and, more importantly, what features make them appear as such, were widely missing. Moreover, the data clearly refuted developers' claim existing robots would be designed in a way not to evoke attributions of gender. Perhaps the prevailing image of robots that was shaped by media (Horstmann & Krämer, 2019; Sandoval et al., 2014) prevents to design and to perceive them as gender-neutral. The commonly shared image of robots seems to determine participants' perceptions of robots and thus likely also participants' willingness to engage in HRI. To investigate possible reasons for participants' hesitations about robots in more detail, a valuable measure of cognitive and affective trust in HRI was provided by Bernotat and colleagues' (2017; 2021) research that was conducted in the context of this research project to prepare the robot stimuli of Experiment 3 and Experiment 4. So far, trust in HRI was considered a unidimensional construct in most prior research. However, the present results imply that the differentiation in cognitive and affective trust in terms of trust in a robot's functions amongst doubts in its motives is needed to deepen insights into participants' perceptions of robots. In this regard, researchers are moreover recommended to consider the actual developmental process of robots that do not come along as humanlike robots as portrayed in media. The present data thus strengthen the call for empirical evidence on participants' perceptions of and associations with various kinds of robots, those that are portrayed in media and those that are actually placed on the market instead of relying on 'gut feelings' and common beliefs (see also Bernotat & Eyssel, 2018). Moreover, as a novelty so far, the present results emphasized the role of social norms and concerns that seem to affect robot perception and thus HRI.

19.8 Conclusion

This PhD project was innovative and ambitious because it combined the fields of social psychology, psycholinguistics, and social robotics. With this valuable unity, existing research questions were answered, new ones were elaborated, and ideas and stimuli for future research developed. This research project therefore provided new insights and perspectives that may enrich scientific discourse and help to keep it ongoing.

References

Abele, A. E. (2003). The dynamics of masculine-agentic and feminine-communal traits: Findings from a prospective study. *Journal of Personality and Social Psychology*, *85*(4), 768–776. https://doi.org/10.1037/0022-3514.85.4.768

Abele, A. E., Hauke, N., Peters, K., Louvet, E., Szymkow, A., & Duan, Y. (2016). Facets of the fundamental content dimensions: Agency with competence and assertiveness – communion with warmth and morality. *Frontiers in Psychology, 7*(1810)*,* 1–17. https://doi.org/10.3389/fpsyg.2016.01810

Abele, A. E., & Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. *Journal of Personality and Social Psychology*, *93*(5), 751–763. https://doi.org/10.1037/0022-3514.93.5.751

Abele, A. E., & Wojciszke, B. (2014). Communal and agentic content in social cognition. In M. Zanna & J. Olson (Eds.), *Advances in experimental social psychology* (Vol. 50, pp. 195–255). Elsevier. https://doi.org/10.1016/B978-0-12-800284-1.00004-7

Ahmad, M. I., Bernotat, J., Lohan, K., & Eyssel, F. (2019). Trust and cognitive load during human-robot interaction. *ArXiv:1909.05160 [Cs]*. http://arxiv.org/abs/1909.05160

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419–439. https://doi.org/10.1006/jmla.1997.2558

Altmann, G. T. M. (2004). Language-mediated eye movements in the absence of a visual world: The 'blank screen paradigm.' *Cognition*, *93*(2), B79–B87. https://doi.org/10.1016/j.cognition.2004.02.005

Altmann, G. T. M. (2011). Language can mediate eye movement control within 100 milliseconds, regardless of whether there is anything to move the eyes to. *Acta Psychologica*, *137*(2), 190–200. https://doi.org/10.1016/j.actpsy.2010.09.009

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1

Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*(4), 502–518. https://doi.org/10.1016/j.jml.2006.12.004

Altmann, G. T. M., & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation. *Cognition*, *111*(1), 55–71. https://doi.org/10.1016/j.cognition.2008.12.005

Arai, M., van Gompel, R. P. G., & Scheepers, C. (2007). Priming ditransitive structures in comprehension. *Cognitive Psychology*, *54*(3), 218–250. https://doi.org/10.1016/j.cogpsych.2006.07.001

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of Other in the Self Scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology*, *63*(4), 596–612. https://doi.org/10.1037/0022-3514.63.4.596

Ashmore, R. D., & Del Boca, F. K. (1979). Sex stereotypes and implicit personality theory: Toward a cognitive—Social psychological conceptualization. *Sex Roles*, *5*(2), 219–248. https://doi.org/10.1007/BF00287932

Athenstaedt, U. (2000). Normative Geschlechtsrollenorientierung: Entwicklung und Validierung eines Fragebogens [Normative Gender Role Orientation: Development and validation of a questionnaire]. *Zeitschrift für Differentielle und Diagnostische Psychologie*, *21*(1), 91–104. https://doi.org/10.1024//0170-1789.21.1.91

Audacity Team. (2015). *Audacity* (Version 2.1.0) [Computer software]. https://audacity.de.uptodown.com/windows/download/187893

Auster, C. J., & Ohm, S. C. (2000). Masculinity and femininity in contemporary American society: A reevaluation using the Bem Sex-Role Inventory. *Sex Roles*, *43*(7), 499–528.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation. Advances in research and theory* (Vol. 8, pp. 47–89). Academic Press. https://doi.org/10.1016/S0079-7421(08)60452-1

Baddeley, A. D., Hitch, G. J., & Allen, R. J. (2009). Working memory and binding in sentence recall. *Journal of Memory and Language*, *61*(3), 438–456. https://doi.org/10.1016/j.jml.2009.05.004

Bakan, D. (1966). *The duality of human existence: An essay on psychology and religion.* Rand Mcnally.

Bargh, J. A. (1999). The cognitive monster: The case against the controllability of automatic stereotype effects. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 361–382). The Guilford Press.

Baron, R. S., Burgess, M. L., & Kao, C. F. (1991). Detecting and labeling prejudice: Do female perpetrators go undetected? *Personality and Social Psychology Bulletin*, *17*(2), 115–123. https://doi.org/10.1177/0146167291172001

Barr, D. J. (2008). Pragmatic expectations and linguistic evidence: Listeners anticipate but do not integrate common ground. *Cognition*, *109*(1), 18–40. https://doi.org/10.1016/j.cognition.2008.07.005

Barreto, M., & Ellemers, N. (2005a). The burden of benevolent sexism: How it contributes to the maintenance of gender inequalities. *European Journal of Social Psychology*, *35*(5), 633–642. https://doi.org/10.1002/ejsp.270

Barreto, M., & Ellemers, N. (2005b). The perils of political correctness: Men's and women's responses to old-fashioned and modern sexist views. *Social Psychology Quarterly*, *68*(1), 75–88. https://doi.org/10.1177/019027250506800106

Barsalou, L. W. (1999). Perceptions of perceptual symbols. *Behavioral and Brain Sciences*, *22*(4), 637–660. https://doi.org/10.1017/S0140525X99532147

Bartneck, C., Yogeeswaran, K., Ser, Q. M., Woodward, G., Sparrow, R., Wang, S., & Eyssel, F. (2018). Robots and racism. In T. Kanda & S. Ŝabanović (Eds.), *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 196–204). Association for Computing Machinery (ACM). https://doi.org/10.1145/3171221.3171260

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2018). Parsimonious mixed models. *ArXiv:1506.04967 [Stat]*. http://arxiv.org/abs/1506.04967

Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology*, *42*(2), 155–162. https://doi.org/10.1037/h0036215

Bem, S. L. (1981). Gender schema theory: A cognitive account of sex typing. *Psychological Review*, *88*(4), 354–364. https://doi.org/10.1037/0033-295X.88.4.354

Berger, A., & Krahé, B. (2013). Negative attributes are gendered too: Conceptualizing and measuring positive and negative facets of sex-role identity: Positive and negative facets of sex-role identity. *European Journal of Social Psychology*, *43*, 516–531. https://doi.org/10.1002/ejsp.1970

Bernotat, J. (2015). *On the tracks of prejudice: Do gender-stereotypical sentences influence gaze directions?* [Unpublished master's thesis]. Bielefeld University.

Bernotat, J., & Eyssel, F. (2018). Can('t) wait to have a robot at home? - Japanese and German users' attitudes toward service robots in smart homes. In J.-J. Cabibihan, F. Mastrogiovanni, A. K. Pandey, S. Rossi, & M. Staffa (Eds.), *Proceedings of the 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (pp. 15–22). IEEE. https://doi.org/10.1109/ROMAN.2018.8525659

Bernotat, J., Eyssel, F., & Sachse, J. (2017). Shape it – The influence of robot body shape on gender perception in robots. In A. Kheddar, E. Yoshida, S. S. Ge, K. Suzuki, J.-J. Cabibihan, F. Eyssel, & H. He (Eds.), *Social Robotics* (Vol. 10652, pp. 75–84). Springer International Publishing. https://doi.org/10.1007/978-3-319-70022-9_8

Bernotat, J., Eyssel, F., & Sachse, J. (2021). The (fe)male robot: How robot body shape impacts first impressions and trust towards robots. *International Journal of Social Robotics*, *13*, 477–489. https://doi.org/10.1007/s12369-019-00562-7

Bernotat, J., Nakanishi, Eyssel, F., & Sumioka. (2021). *Give me a hug – How 'Hugvie', a huggable communication medium, affects impression formation* [Manuscript in preparation]. Department of Psychology, Bielefeld University.

Bernotat, J., Schiffhauer, B., Eyssel, F., Holthaus, P., Leichsenring, C., Richter, V., Pohling, M., Carlmeyer, B., Köster, N., Meyer zu Borgsen, S., Zorn, R., Engelmann, K. F., Lier, F., Schulz, S., Bröhl, R., Seibel, E., Hellwig, P., Cimiano, P., Kummert, F., … Wrede, S. (2016). Welcome to the future – How naïve users intuitively address an intelligent robotics apartment. In A. Agah, J.-J. Cabibihan, A. M. Howard, M. A. Salichs, & H. He (Eds.), *Social Robotics* (Vol. 9979, pp. 982–992). Springer International Publishing. https://doi.org/10.1007/978-3-319-47437-3_96

Blair, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review*, *6*(3), 242–261. https://doi.org/10.1207/S15327957PSPR0603_8

Bluemke, M., & Friese, M. (2006). Do features of stimuli influence IAT effects? *Journal of Experimental Social Psychology*, *42*(2), 163–176. https://doi.org/10.1016/j.jesp.2005.03.004

Bodenhausen, G. V., Kang, S. K., & Peery, D. (2012). Social categorization and the perception of social groups. In S. T. Fiske & C. N. Macrae (Eds.), *The SAGE handbook of social cognition* (pp. 311–329). SAGE Publications Ltd. https://doi.org/10.4135/9781446247631.n16

Boninger, D. S., Krosnick, J. A., & Berent, M. K. (1995). Origins of attitude importance: Self-interest, social identification, and value relevance. *Journal of Personality and Social Psychology*, *68*(1), 61–80. https://doi.org/10.1037/0022-3514.68.1.61

Bourassa, D. C., McManus, I. C., & Bryden, M. P. (1996). Handedness and eye-dominance: A meta-analysis of their relationship. *Laterality*, *1*(1), 5-34. https://doi.org/10.1080/713754206

Bransford, J. D., Barclay, J. R., & Franks, J. J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology*, *3*(2), 193–209. https://doi.org/10.1016/0010-0285(72)90003-5

Breazeal, C. (2003). Emotion and sociable humanoid robots. *International Journal of Human-Computer Studies*, *59*(1–2), 119–155. https://doi.org/10.1016/S1071-5819(03)00018-1

Bruckenberger, U., Weiss, A., Mirnig, N., Strasser, E., Stadler, S., & Tscheligi, M. (2013). The good, the bad, the weird: Audience evaluation of a “real” robot in relation to science fiction and mass media. In G. Herrmann, M. J. Pearson, A. Lenz, P. Bremner, A. Spiers, & U. Leonards (Eds.), *Social Robotics* (Vol. 8239, pp. 301–310). Springer International Publishing. https://doi.org/10.1007/978-3-319-02675-6_30

Calvert, L. M., & Ramsey, V. J. (1996). Speaking as female and white: A non-dominant/dominant group standpoint. *Organization*, *3*(4), 468–485. https://doi.org/10.1177/135050849634004

Campana, G., & Casco, C. (2009). Repetition effects of features and spatial position: Evidence for dissociable mechanisms. *Spatial Vision*, *22*(4), 325–338. https://doi.org/10.1163/156856809788746318

Canal, P., Garnham, A., & Oakhill, J. (2015). Beyond gender stereotypes in language comprehension: Self sex-role descriptions affect the brain’s potentials associated with agreement processing. *Frontiers in Psychology*, *6*(1953), 1-17. https://doi.org/10.3389/fpsyg.2015.01953

Čapek, K. (1921). *R.U.R. - Rossum Universal Robots*. Wildside Press.

Carpenter, J., Davis, J. M., Erwin-Stewart, N., Lee, T. R., Bransford, J. D., & Vye, N. (2009). Gender representation and humanoid robots designed for domestic use. *International Journal of Social Robotics*, *1*(3), 261–265. https://doi.org/10.1007/s12369-009-0016-4

Carreiras, M., Garnham, A., Oakhill, J., & Cain, K. (1996). The use of stereotypical gender information in constructing a mental model: Evidence from English and Spanish. *The Quarterly Journal of Experimental Psychology Section A*, *49*(3), 639–663. https://doi.org/10.1080/713755647

Castillo-Mayén, R., & Montes-Berges, B. (2017). Which automatic associations prevail? Congruency and reverse priming effects on implicit gender stereotyping. *Pensamiento Psicológico*, *15*(1), 33–49.

Cedrus Corporation. (2021). *RB-844 Response Pad* [Apparatus]. https://cedrus.com/store/rb844.html

Chen, M., & Bargh, J. A. (1997). Nonconscious behavioral confirmation processes: The self-fulfilling consequences of automatic stereotype activation. *Journal of Experimental Social Psychology*, *33*(5), 541–560. https://doi.org/10.1006/jesp.1997.1329

Chen, M., & Bargh, J. A. (1999). Consequences of automatic evaluation: Immediate behavioral predispositions to approach or avoid the stimulus. *Personality and Social Psychology Bulletin*, *25*(2), 215–224. https://doi.org/10.1177/0146167299025002007

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*(4), 335–359. https://doi.org/10.1016/S0022-5371(73)80014-3

Clark, D. M., Ball, S., & Pape, D. (1991). An experimental investigation of thought suppression. *Behaviour Research and Therapy*, *29*(3), 253–257. https://doi.org/10.1016/0005-7967(91)90115-J

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155-159.

Connine, C. M., Mullennix, J., Shernoff, E., & Yelen, J. (1990). Word familiarity and frequency in visual and auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(6), 1084–1096. https://doi.org/10.1037/0278-7393.16.6.1084

Connor, C. E., Egeth, H. E., & Yantis, S. (2004). Visual attention: Bottom-up versus top-down. *Current Biology*, *14*(19), R850–R852. https://doi.org/10.1016/j.cub.2004.09.041

Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. J. (2005). Separating multiple processes in implicit social cognition: The quad model of implicit task performance. *Journal of Personality and Social Psychology*, *89*(4), 469–487. https://doi.org/10.1037/0022-3514.89.4.469

Contreras, J. M., Banaji, M. R., & Mitchell, J. P. (2012). Dissociable neural correlates of stereotypes and other forms of semantic knowledge. *Social Cognitive and Affective Neuroscience*, *7*(7), 764–770. https://doi.org/10.1093/scan/nsr053

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language. *Cognitive Psychology*, *6*(1), 84–107. https://doi.org/10.1016/0010-0285(74)90005-X

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*(1), 98–104. https://doi.org/10.1037/0021-9010.78.1.98

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334. https://doi.org/10.1007/BF02310555

Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: Immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(2), 498–513. https://doi.org/10.1037/0278-7393.30.2.498

Dahan, D., & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletin & Review*, *12*(3), 453–459. https://doi.org/10.3758/BF03193787

Dang, J., & Liu, L. (2021). Robots are friends as well as foes: Ambivalent attitudes toward mindful and mindless AI robots in the United States and China. *Computers in Human Behavior*, *115*, 106612-106648. https://doi.org/10.1016/j.chb.2020.106612

Davison, G. C., Navarre, S. G., & Vogel, R. S. (1995). The articulated thoughts in simulated situations paradigm: A think-aloud approach to cognitive assessment. *Current Directions in Psychological Science*, *4*(1), 29–33. https://doi.org/10.1111/1467-8721.ep10770963

De Bernardi, J. (1994). Social aspects of language use. In T. Ingold (Ed.), *Companion encyclopedia of anthropology* (pp. 861–890). Routledge. http://www.sub.uni-hamburg.de/ebook/ebook.php?act=b&cid=2277

De Deyne, S., Cabana, Á., Li, B., Cai, Q., & McKague, M. (2020). A cross-linguistic study into the contribution of affective connotation in the lexico-semantic representation of concrete and abstract concepts. In S. Denison, M. Mack, Y. Xu, & B. C. Armstrong (Eds.), *Proceedings of the 42nd Annual Conference of the Cognitive Science Society (CogSci)* (pp. 2776–2782). Cognitive Science Society.

de Graaf, M. M. A., & Allouch, S. B. (2013). The relation between people's attitude and anxiety towards robots in human-robot interaction. *Proceedings of the 22th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (pp. 632–637). https://doi.org/10.1109/ROMAN.2013.6628419

De Houwer, J. (2009). The propositional approach to associative learning as an alternative for association formation models. *Learning & Behavior*, *37*(1), 1–20. https://doi.org/10.3758/LB.37.1.1

De Judicibus, M., & McCabe, M. P. (2001). Blaming the target of sexual harassment: Impact of gender role, sexist attitudes, and work role. *Sex Roles*, *44*(7-8), 401–417. https://doi.org/10.1023/A:1011926027920

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*(1), 5–18. https://doi.org/10.1037/0022-3514.56.1.5

Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems*, *42*(3-4), 177–190. https://doi.org/10.1016/S0921-8890(02)00374-3

Duffy, S. A., & Keir, J. A. (2004). Violating stereotypes: Eye movements and comprehension processes when text conflicts with world knowledge. *Memory & Cognition*, *32*(4), 551–559. https://doi.org/10.3758/BF03195846

Duñabeitia, J. A., Avilés, A., Afonso, O., Scheepers, C., & Carreiras, M. (2009). Qualitative differences in the representation of abstract versus concrete words: Evidence from the visual world paradigm. *Cognition*, *110*(2), 284–292. https://doi.org/10.1016/j.cognition.2008.11.012

Eagly, A. H., & Mladinic, A. (1989). Gender stereotypes and attitudes toward women and men. *Personality and Social Psychology Bulletin*, *15*(4), 543–558. https://doi.org/10.1177/0146167289154008

Eagly, A. H., Nater, C., Miller, D. I., Kaufmann, M., & Sczesny, S. (2019). Gender stereotypes have changed: A cross-temporal meta-analysis of U.S. public opinion polls from 1946 to 2018. *American Psychologist*, *75*(3), 301–315. https://doi.org/10.1037/amp0000494

Eckes, T., & Six-Materna, I. (1999). Hostilität und Benevolenz: Eine Skala zur Erfassung des ambivalenten Sexismus [Hostility and benevolence: A scale to measure ambivalent sexism]. *Zeitschrift für Sozialpsychologie*, *30*(4), 211–228. https://doi.org/10.1024//0044-3514.30.4.211

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, *114*(4), 864–886. https://doi.org/10.1037/0033-295X.114.4.864

Erdogan, A. R., Özdikici, M., Aydin, M. D., Aktas, Ö., & Dane, S. (2002). Right and left visual cortex areas in healthy subjects with right- and left-eye dominance. *International Journal of Neuroscience*, *112*(5), 517–523. https://doi.org/10.1080/00207450290025626

Esaulova, Y., Reali, C., & von Stockhausen, L. (2014). Influences of grammatical and stereotypical gender during reading: Eye movements in pronominal and noun phrase anaphor resolution. *Language, Cognition and Neuroscience*, *29*(7), 781–803. https://doi.org/10.1080/01690965.2013.794295

Eyssel, F. (2010, September 26–30). *Neuentwicklung und Validierung einer deutschsprachigen Skala zur internalen und externalen Kontrolle von sexistischem Verhalten*. [Conference session]. 47. Kongress der Deutschen Gesellschaft für Psychologie (DGPs), Bremen, Germany. https://www.dgps.de/index.php?id=2000501

Eyssel, F., & Hegel, F. (2012). (S)he's got the look: Gender stereotyping of robots. *Journal of Applied Social Psychology*, *42*(9), 2213–2230. https://doi.org/10.1111/j.1559-1816.2012.00937.x

Eyssel, F., & Kuchenbrandt, D. (2012). Social categorization of social robots: Anthropomorphism as a function of robot group membership. *British Journal of Social Psychology*, *51*(4), 724–731. https://doi.org/10.1111/j.2044-8309.2011.02082.x

Eyssel, F., Kuchenbrandt, D., & Bobinger, S. (2011). Effects of anticipated human-robot interaction and predictability of robot behavior on perceptions of anthropomorphism. In H. Cramer (Ed.), *Proceedings of the 6th international conference on Human-robot interaction (HRI)* (pp. 61-68). Association for Computing Machinery (ACM).

Eyssel, F., & Loughnan, S. (2013). "It don't matter if you're black or white"? In G. Herrmann, M. J. Pearson, A. Lenz, P. Bremner, A. Spiers, & U. Leonards (Eds.), *Social Sobotics* (Vol. 8239, pp. 422–431). Springer International Publishing. https://doi.org/10.1007/978-3-319-02675-6_42

Ezer, N. (2008). *Is a robot an appliance, teammate, or friend? Age-related differences in expectations of and attitudes towards personal home-based robots* (Publication No. 3345975) [Doctoral dissertation, Georgia Institute of Technology]. ProQuest Dissertations Publishing.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*,*41*, 1149-1160.

Faul, F., Erdfelder, E., Lang, A.-G., Buchner, A. (2007). *G*Power* (Version 3.1) [Statistic software]. https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower

Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The mode model as an integrative framework. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 75–109). Academic Press. https://doi.org/10.1016/S0065-2601(08)60318-4

Fazio, R. H. (2007). Attitudes as object–evaluation associations of varying strength. *Social Cognition*, *25*(5), 603–637. https://doi.org/10.1521/soco.2007.25.5.603

Fazio, R. H., Chen, J., McDonel, E. C., & Sherman, S. J. (1982). Attitude accessibility, attitude-behavior consistency, and the strength of the object-evaluation association. *Journal of Experimental Social Psychology*, *18*(4), 339–357. https://doi.org/10.1016/0022-1031(82)90058-0

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology*, *69*(6), 1013–1027. https://doi.org/10.1037/0022-3514.69.6.1013

Fernandes, E., Costa, A., & Coco, M. I. (2015). Bridging mechanisms of reading, viewing and working memory during attachment resolution of ambiguous relative clauses. In G. Airenti, B. G. Bara, G. Sandini, & M. Cruciani (Eds.), *Proceedings of the Euro Asian Pacific Joint Conference on Cognitive Science / 4th European Conference on Cognitive Science / 11th International Conference on Cognitive Science,* (Vol. 1419, pp. 264–269). CEUR-WS.org. http://ceur-ws.org/Vol-1419/paper0040.pdf

Fiedler, K., Messner, C., & Bluemke, M. (2006). Unresolved problems with the "I", the "A", and the "T": A logical and psychometric critique of the Implicit Association Test (IAT). *European Review of Social Psychology*, *17*(1), 74–147. https://doi.org/10.1080/10463280600681248

Field, A. P. (2013). *Discovering statistics using SPSS* (4th ed). SAGE Publications.

Finnegan, E., Garnham, A., & Oakhill, J. (2015). Social consensus feedback as a strategy to overcome spontaneous gender stereotypes. *Discourse Processes*, *52*(5–6), 434–462. https://doi.org/10.1080/0163853X.2015.1026680

Fiske, S. T. (1998). Stereotyping, prejudice, and discrimination. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (pp. 357–411). McGraw-Hill.

Fiske, S. T. (2018). Stereotype content: Warmth and competence endure. *Current Directions in Psychological Science*, *27*(2), 67–73. https://doi.org/10.1177/0963721417738825

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, *11*(2), 77–83. https://doi.org/10.1016/j.tics.2006.11.005

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878–902. https://doi.org/10.1037/0022-3514.82.6.878

Fiske, S. T., Xu, J., Cuddy, A. C., & Glick, P. (1999). (Dis)respecting versus (dis)liking: Status and interdependence predict ambivalent stereotypes of competence and warmth. *Journal of Social Issues*, *55*(3), 473–489.

Frennert, S., Eftring, H., & Östlund, B. (2013). What older people expect of robots: A mixed methods approach. In G. Herrmann, M. J. Pearson, A. Lenz, P. Bremner, A. Spiers, & U. Leonards (Eds.), *Social Robotics* (Vol. 8239, pp. 19–29). Springer International Publishing. https://doi.org/10.1007/978-3-319-02675-6_3

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological Bulletin*, *133*(4), 694–724. https://doi.org/10.1037/0033-2909.133.4.694

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, *141*(1), 2–18. https://doi.org/10.1037/a0024338

Gabriel, U., Behne, D. M., & Gygax, P. M. (2017). Speech vs. reading comprehension: An explorative study of gender representations in Norwegian. *Journal of Cognitive Psychology*, *29*(7), 795–808. https://doi.org/10.1080/20445911.2017.1326923

Garnham, A. (1981). Mental models as representations of text. *Memory & Cognition*, *9*(6), 560–565. https://doi.org/10.3758/BF03202350

Garnham, A., Oakhill, J., & Reynolds, D. (2002). Are inferences from stereotyped role names to characters' gender made elaboratively? *Memory & Cognition*, *30*(3), 439–446. https://doi.org/10.3758/BF03194944

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*(5), 692–731. https://doi.org/10.1037/0033-2909.132.5.692

Gawronski, B., & Bodenhausen, G. V. (2011). The associative–propositional evaluation model. In M. P. Zanna & J. M. Olson (Eds.), *Advances in experimental social psychology* (Vol. 44, pp. 59–127). Elsevier. https://doi.org/10.1016/B978-0-12-385522-0.00002-0

Gawronski, B., & De Houwer, J. (2014). Implicit measures in social and personality psychology. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (pp. 283–310). Cambridge University Press.

Gawronski, B., & Hahn, A. (2019). Implicit measures: Procedures, use, and interpretation. In H. Blanton, J. M. LaCroix, & G. D. Webster (Eds.), *Measurement in social psychology* (pp. 29–55). Routledge/Taylor & Francis Group.

Geyer, T., Müller, H. J., & Krummenacher, J. (2007). Cross-trial priming of element positions in visual pop-out search is dependent on stimulus arrangement. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(4), 788–797. https://doi.org/10.1037/0096-1523.33.4.788

Giles, H., Coupland, J., & Coupland, N. (1991). *Contexts of accommodation: Developments in applied sociolinguistics* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511663673

Glick, P., & Fiske, S. T. (1996). The Ambivalent Sexism Inventory: Differentiating hostile and benevolent sexism. *Journal of Personality and Social Psychology*, *70*(3), 491–512. https://doi.org/10.1037/0022-3514.70.3.491

Glick, P., & Fiske, S. T. (2001). An ambivalent alliance: Hostile and benevolent sexism as complementary justifications for gender inequality. *American Psychologist*, *56*(2), 109–118. https://doi.org/10.1037/0003-066X.56.2.109

Goldberg, J. H., & Wichansky, A. M. (2003). Eye tracking in usability evaluation: A practitioner's guide. In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research*. North-Holland.

Govan, C. L., & Williams, K. D. (2004). Changing the affective valence of the stimulus items influences the IAT by re-defining the category labels. *Journal of Experimental Social Psychology*, *40*(3), 357–365. https://doi.org/10.1016/j.jesp.2003.07.002

Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition*, *125*(1), 125–130. https://doi.org/10.1016/j.cognition.2012.06.007

Greenwald, A. G., Klinger, M. R., & Schuh, E. S. (1995). Activation by marginally perceptible ("subliminal") stimuli: Dissociation of unconscious from conscious cognition. *Journal of Experimental Psychology: General*, *124*(1), 22–42. https://doi.org/10.1037/0096-3445.124.1.22

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480.

Guerra, E., Bernotat, J., Carvacho, H., & Bohner, G. (2021). Ladies first: Gender stereotypes drive anticipatory eye-movements during incremental sentence interpretation. *Frontiers in Psychology*, *12*(589429), 1–10. https://doi.org/10.3389/fpsyg.2021.589429

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., de Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *53*(5), 517–527. https://doi.org/10.1177/0018720811417254

Hancock, P. A., Kessler, T. T., Kaplan, A. D., Brill, J. C., & Szalma, J. L. (2020). Evolving trust in robots: Specification through sequential and comparative meta-analyses. *Human Factors: The Journal of the Human Factors and Ergonomics Society,* 1–34. https://doi.org/10.1177/0018720820922080

Haslam, N. (2006). Dehumanization: An integrative review. *Personality and Social Psychology Review*, *10*(3), 252–264. https://doi.org/10.1207/s15327957pspr1003_4

Hauk, O., & Pulvermüller, F. (2004). Effects of word length and frequency on the human event-related potential. *Clinical Neurophysiology*, *115*(5), 1090–1103. https://doi.org/10.1016/j.clinph.2003.12.020

Heilman, M. E. (2012). Gender stereotypes and workplace bias. *Research in Organizational Behavior*, *32*, 113–135. https://doi.org/10.1016/j.riob.2012.11.003

Hewstone, M., & Greenland, K. (2000). Intergroup conflict. *International Journal of Psychology*, *35*(2), 136–144. https://doi.org/10.1080/002075900399439

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*(11), 1539–1558. https://doi.org/10.1002/sim.1186

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ*, *327*(7414), 557–560. https://doi.org/10.1136/bmj.327.7414.557

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.

Hooper & McHugh. (2013). The effects of repeated thought suppression attempts on thought occurrence. *The American Journal of Psychology*, *126*(3), 315. https://doi.org/10.5406/amerjpsyc.126.3.0315

Horstmann, A. C., & Krämer, N. C. (2019). Great expectations? Relation of previous experiences with social robots in real life or in the media and expectancies based on qualitative and quantitative assessment. *Frontiers in Psychology*, *10*(939), 1–14. https://doi.org/10.3389/fpsyg.2019.00939

Huettig, F., & Altmann, G. T. M. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, *96*(1), B23–B32. https://doi.org/10.1016/j.cognition.2004.10.003

Huettig, F., & Altmann, G. T. M. (2007). Visual-shape competition during language-mediated attention is based on lexical input and not modulated by contextual appropriateness. *Visual Cognition*, *15*(8), 985–1018. https://doi.org/10.1080/13506280601130875

Huettig, F., & Altmann, G. T. M. (2011). Looking at anything that is green when hearing “frog”: How object surface colour and stored object colour knowledge influence language-mediated overt attention. *Quarterly Journal of Experimental Psychology*, *64*(1), 122–145. https://doi.org/10.1080/17470218.2010.481474

Huettig, F., & Janse, E. (2016). Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Language, Cognition and Neuroscience*, *31*(1), 80–93. https://doi.org/10.1080/23273798.2015.1047459

Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language*, *57*(4), 460–482. https://doi.org/10.1016/j.jml.2007.02.001

Huettig, F., Olivers, C. N. L., & Hartsuiker, R. J. (2011a). Looking, language, and memory: Bridging research from the visual world and visual search paradigms. *Acta Psychologica*, *137*(2), 138–150. https://doi.org/10.1016/j.actpsy.2010.07.013

Huettig, F., Rommers, J., & Meyer, A. S. (2011b). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171. https://doi.org/10.1016/j.actpsy.2010.11.003

Huettig, F., Singh, N., & Mishra, R. K. (2011c). Language-mediated visual orienting behavior in low and high literates. *Frontiers in Psychology*, *2*(285), 1–14. https://doi.org/10.3389/fpsyg.2011.00285

Hyönä, J., & Olson, R. K. (1995). Eye fixation patterns among dyslexic and normal readers: Effects of word length and word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(6), 1430–1440. https://doi.org/10.1037/0278-7393.21.6.1430

Inhoff, A. W., & Radach, R. (1998). Definition and computation of oculomotor measures in the study of cognitive processes. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 29–53). Elsevier. https://doi.org/10.1016/B978-008043361-5/50003-1

Irmen, L. (2007). What’s in a (role) name? Formal and conceptual aspects of comprehending personal nouns. *Journal of Psycholinguistic Research*, *36*(6), 431–456. https://doi.org/10.1007/s10936-007-9053-z

Irmen, L., & Roßberg, N. (2004). Gender markedness of language: The impact of grammatical and nonlinguistic information on the mental representation of person information. *Journal of Language and Social Psychology*, *23*(3), 272–307. https://doi.org/10.1177/0261927X04266810

Irmen, L., & Roßberg, N. (2006). How formal versus semantic gender influences the interpretation of person denotations. *Swiss Journal of Psychology*, *65*(3), 157–165. https://doi.org/10.1024/1421-0185.65.3.157

Jackson, R. B., & Williams, T. (2019). *On perceived social and moral agency in natural language capable robots*. In H. W. Park (Ed.), *Proceedings of the 2019 ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 401–410). Association for Computing Machinery (ACM).

Johnson-Laird, P. N. (1980). Mental models in cognitive science. *Cognitive Science*, *4*(1), 71–115. https://doi.org/10.1207/s15516709cog0401_4

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge, MA: Harvard University Press. https://hal.archives-ouvertes.fr/hal-00702919

Jung, E. H., Waddell, T. F., & Sundar, S. S. (2016). Feminizing robots: User responses to gender cues on robot body and screen. In J. Kayne & A. Druin, *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 3107–3113. Association for Computing Machinery (ACM). https://doi.org/10.1145/2851581.2892428

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, *8*(4), 441–480. https://doi.org/10.1016/0010-0285(76)90015-3

Just, M. A., & Carpenter, P. A. (1993). The intensity dimension of thought: Pupillometric indices of sentence processing. *Canadian Journal of Experimental Psychology/Revue Canadienne de psychologie expérimentale,* *47*(2), 310–339. https://doi.org/10.1037/h0078820

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*(1), 133–156. https://doi.org/10.1016/S0749-596X(03)00023-8

Kennison, S. M., & Trofe, J. L. (2003). Comprehending pronouns: A role for word-specific gender stereotype information. *Journal of Psycholinguistic Research*, *32*(3), 355–378. https://doi.org/10.1023/A:1023599719948

Klonis, S. C., Plant, E. A., & Devine, P. G. (2005). Internal and external motivation to respond without sexism. *Personality and Social Psychology Bulletin*, *31*(9), 1237–1249. https://doi.org/10.1177/0146167205275304

Knoeferle, P., & Crocker, M. W. (2006). The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye tracking. *Cognitive Science*, *30*(3), 481–529. https://doi.org/10.1207/s15516709cog0000_65

Knoeferle, P., & Crocker, M. W. (2007). The influence of recent scene events on spoken comprehension: Evidence from eye movements. *Journal of Memory and Language*, *57*(4), 519–543. https://doi.org/10.1016/j.jml.2007.01.003

Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition*, *95*(1), 95–127. https://doi.org/10.1016/j.cognition.2004.03.002

Knoeferle, P., Urbach, T. P., & Kutas, M. (2014). Different mechanisms for role relations versus verb–action congruence effects: Evidence from ERPs in picture–sentence verification. *Acta Psychologica*, *152*, 133–148. https://doi.org/10.1016/j.actpsy.2014.08.004

Koch, C., & Tsuchiya, N. (2007). Attention and consciousness: Two distinct brain processes. *Trends in Cognitive Sciences*, *11*(1), 16–22. https://doi.org/10.1016/j.tics.2006.10.012

Koenig, A. M., & Eagly, A. H. (2014). Evidence for the social role theory of stereotype content: Observations of groups' roles shape stereotypes. *Journal of Personality and Social Psychology*, *107*(3), 371–392. https://doi.org/10.1037/a0037215

Kröger, J. M., Münster, K., Burigo, M., & Knoeferle, P. (2018). Wiggle, wiggle, wiggle: How visual cues influence thematic role assignment in children and adults. In T. T. Rogers, M. Rau, X. Zhu, & C. W. Kalish (Eds.), *Proceedings of the 40th Annual Conference of the Cognitive Science Society (CogSci)* (pp. 1969–1974). Cognitive Science Society.

Labov, W. (1972). Some principles of linguistic methodology. *Language in Society*, *1*(1), 97–120. https://doi.org/10.1017/S0047404500006576

Lepore, S. J., Allen, K. A., & Evans, G. W. (1993). Social support lowers cardiovascular reactivity to an acute stressor. *Psychosomatic Medicine*, *55*(6), 518–524. https://doi.org/10.1097/00006842-199311000-00007

Loughnan, S., & Haslam, N. (2007). Animals and androids: Implicit associations between social categories and nonhumans. *Psychological Science*, *18*(2), 116–121. https://doi.org/10.1111/j.1467-9280.2007.01858.x

Lowery, B. S., Hardin, C. D., & Sinclair, S. (2001). Social influence effects on automatic racial prejudice. *Journal of Personality and Social Psychology*, *81*(5), 842–855. https://doi.org/10.1037/0022-3514.81.5.842

MacDorman, K. F., Vasudevan, S. K., & Ho, C.-C. (2009). Does Japan really have robot mania? Comparing attitudes by implicit and explicit measures. *AI & Society*, *23*(4), 485–510. https://doi.org/10.1007/s00146-008-0181-2

Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (1999). Spoken word recognition in the visual world paradigm reflects the structure of the entire lexicon. In M. Hahn & S. C. Stoness (Eds.), *Proceedings of the 21st annual meeting of the Cognitive Science Society* (pp. 331–336). Psychology Press.

Makatchev, M., Simmons, R., Sakr, M., & Ziadee, M. (2013). Expressing ethnicity through behaviors of a robot character. In H. Kuzuoka, V. Evers, M. Imai, & J. Forlizzi (Eds.), *Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 357–364). IEEE. https://doi.org/10.1109/HRI.2013.6483610

Maljkovic, V., & Nakayama, K. (1996). Priming of pop-out: II. The role of position. *Perception & Psychophysics*, *58*(7), 977–991. https://doi.org/10.3758/BF03206826

Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccades. *Perception & Psychophysics*, *53*(4), 372–380. https://doi.org/10.3758/BF03206780

McQueen, J. M., & Huettig, F. (2012). Changing only the probability that spoken words will be distorted changes how they are recognized. *The Journal of the Acoustical Society of America*, *131*(1), 509–517. https://doi.org/10.1121/1.3664087

McShane, B. B., & Böckenholt, U. (2017). Single paper meta-analysis: Benefits for study summary, theory-testing, and replicability. *Journal of Consumer Research*, 1048–1063. https://doi.org/10.1093/jcr/ucw085

Mendes, W. B., Blascovich, J., Lickel, B., & Hunter, S. (2002). Challenge and threat during social interactions with white and black men. *Personality and Social Psychology Bulletin*, *28*(7), 939–952. https://doi.org/10.1177/01467202028007007

Minadakis, G., & Lohan, K. (2018). Using pupil diameter to measure cognitive load. *ArXiv:1812.07653 [Cs]*. http://arxiv.org/abs/1812.07653

Mitchell, J. P., Nosek, B. A., & Banaji, M. R. (2003). Contextual variations in implicit evaluation. *Journal of Experimental Psychology: General*, *132*(3), 455–469. https://doi.org/10.1037/0096-3445.132.3.455

Molinaro, N., Su, J.-J., & Carreiras, M. (2016). Stereotypes override grammar: Social knowledge in sentence comprehension. *Brain and Language*, *155–156*, 36–43. https://doi.org/10.1016/j.bandl.2016.03.002

Monteith, M. J. (1993). Self-regulation of prejudiced responses: Implications for progress in prejudice-reduction efforts. *Journal of Personality and Social Psychology*, *65*(3), 469–485. https://doi.org/10.1037/0022-3514.65.3.469

Monteith, M. J., Ashburn-Nardo, L., Voils, C. I., & Czopp, A. M. (2002). Putting the brakes on prejudice: On the development and operation of cues for control. *Journal of Personality and Social Psychology*, *83*(5), 1029–1050. https://doi.org/10.1037/0022-3514.83.5.1029

Moreno, I., Coco, M. I., & Keller, F. (2015). The interaction of visual and linguistic saliency during syntactic ambiguity resolution. *Quarterly Journal of Experimental Psychology*, *68*(1), 46–74. https://doi.org/10.1080/17470218.2014.936475

Mori, M. (1970). Bukimi no tani [The uncanny valley]. *Energy*, *7*, 33–35.

Morris, P. E., & Fritz, C. O. (2013). Effect sizes in memory research. *Memory*, *21*(7), 832–842. https://doi.org/10.1080/09658211.2013.763984

Müller, C. (2020, September 24). *IFR Press Conference.* International Federation of Social Robotics (IFR). https://ifr.org/downloads/press2018/Presentation_WR_2020.pdf

Münster, K., & Knoeferle, P. (2017). Situated language processing across the lifespan: A review. *International Journal of English Linguistics*, *7*(1), 1–13. https://doi.org/10.5539/ijel.v7n1p1

Münster, K., & Knoeferle, P. (2018). Extending situated language comprehension (accounts) with speaker and comprehender characteristics: Toward socially situated interpretation. *Frontiers in Psychology*, *8*(2267), 1–12. https://doi.org/10.3389/fpsyg.2017.02267

Neyer, F. J., Felber, J., & Gebhardt, C. (2012). Entwicklung und Validierung einer Kurzskala zur Erfassung von Technikbereitschaft [Development and validation of a short scale to measure technology commitment]. *Diagnostica*, *58*(2), 87–99. https://doi.org/10.1026/0012-1924/a000067

Niederhoffer, K. G., & Pennebaker, J. W. (2002). Linguistic style matching in social interaction. *Journal of Language and Social Psychology*, *21*(4), 337–360. https://doi.org/10.1177/026192702237953

Norris, C. J., Chen, E. E., Zhu, D. C., Small, S. L., & Cacioppo, J. T. (2004). The interaction of social and emotional processes in the brain. *Journal of Cognitive Neuroscience*, *16*(10), 1818–1829. https://doi.org/10.1162/0898929042947847

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed). McGraw-Hill.

Osterhout, L., Bersick, M., & Mclaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Memory & Cognition*, *25*(3), 273–285. https://doi.org/10.3758/BF03211283

Otterbacher, J., & Talias, M. (2017). S/he's too warm/agentic!: The influence of gender on uncanny reactions to robots. *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, 214–223. https://doi.org/10.1145/2909824.3020220

Pettijohn, T. F. I., & Walzer, A. S. (2008). Reducing racism, sexism, and homophobia in college students by completing a psychology of prejudice course. *College Student Journal*, *42*(2), 459–468.

Phelan, J. E., & Rudman, L. A. (2010). Prejudice toward female leaders: Backlash effects and women's impression management dilemma: Prejudice toward female leaders. *Social and Personality Psychology Compass*, *4*(10), 807–820. https://doi.org/10.1111/j.1751-9004.2010.00306.x

Plant, E. A., & Devine, P. G. (2009). The active control of prejudice: Unpacking the intentions guiding control efforts. *Journal of Personality and Social Psychology*, *96*(3), 640–652. https://doi.org/10.1037/a0012960

Pollatsek, A., & Well, A. D. (1995). On the use of counterbalanced designs in cognitive research: A suggestion for a better and more powerful analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(3), 785–794. https://doi.org/10.1037/0278-7393.21.3.785

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3–25. https://doi.org/10.1080/00335558008248231

Powers, A., Kramer, A. D. I., Lim, S., Kuo, J., Lee, Sau-lai, & Kiesler, S. (2005). Eliciting information from people with a gendered humanoid robot. *Proceedings of the 18th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (pp. 158–163). IEEE. https://doi.org/10.1109/ROMAN.2005.1513773

Průcha, J. (1972). Psycholinguistics and sociolinguistics—Separate or integrated? *Linguistics*, *10*(89). https://doi.org/10.1515/ling.1972.10.89.9

Pyykkönen, P., Hyönä, J., & van Gompel, R. P. G. (2010). Activating gender stereotypes during online spoken language processing: Evidence from visual world eye tracking. *Experimental Psychology*, *57*(2), 126–133. https://doi.org/10.1027/1618-3169/a000016

Qualtrics^{XM}. (2018). *Qualtrics* [Survey software]. https://www.qualtrics.com/de/

Questback. (2017). *Unipark* [Survey software]. https://www.unipark.com/

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. https://doi.org/10.1037/0033-2909.124.3.372

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, *14*(3), 191–201. https://doi.org/10.3758/BF03197692

Reich, N., & Eyssel, F. (2013). Attitudes towards service robots in domestic environments: The role of personality characteristics, individual interests, and demographic variables. *Paladyn, Journal of Behavioral Robotics*, *4*(2), 123–130. https://doi.org/10.2478/pjbr-2013-0014

Reich-Stiebert, N., & Eyssel, F. (2015). Learning with educational companion robots? Toward attitudes on education robots, predictors of attitudes, and application potentials for education robots. *International Journal of Social Robotics*, *7*(5), 875–888. https://doi.org/10.1007/s12369-015-0308-9

Reich-Stiebert, N., & Eyssel, F. (2016). Robots in the classroom: What teachers think about teaching and learning with education robots. In A. Agah, J.-J. Cabibihan, A. M. Howard, M. A. Salichs, & H. He (Eds.), *Social Robotics* (Vol. 9979, pp. 671–680). Springer International Publishing. https://doi.org/10.1007/978-3-319-47437-3_66

Ren, P., Barreto, A., Huang, J., Gao, Y., Ortega, F. R., & Adjouadi, M. (2014). Off-line and online stress detection through processing of the pupil diameter signal. *Annals of Biomedical Engineering*, *42*(1), 162–176. https://doi.org/10.1007/s10439-013-0880-9

Reysen, S. (2005). Construction of a new scale: The Reysen Likability Scale. *Social Behavior and Personality: An International Journal*, *33*(2), 201–208. https://doi.org/10.2224/sbp.2005.33.2.201

Rickheit, G., Sichelschmidt, L., & Strohner, H. (2004). *Psycholinguistik: Die Wissenschaft vom sprachlichen Verhalten und Erleben* [Psycholinguistics: The science of linguistic behavior and perception] (2nd ed.). Stauffenburg Verlag.

Rodriguez, A., Burigo, M., & Knoeferle, P. (2015). Visual gender cues elicit agent expectations: Different mismatches in situated language comprehension. In G. Airenti, B. G. Bara, & G. Sandini (Eds.), *Proceedings of the EuroAsianPacific Joint Conference on Cognitive Science (EAPcogsci 2015)* (Vol. 1419, pp. 234–239).

Rodriguez, A., Burigo, M., & Knoeferle, P. (2016). Visual constraints modulate stereotypical predictability of agents during situated language comprehension. In A. Papafragou, D. J. Grodner, D. Mirman, & J. Trueswell (Eds.), *Proceedings of the 38th Annual Cognitive Science Society Meeting*. https://nbn-resolving.org/urn:nbn:de:0070-pub-29027421, https://pub.uni-bielefeld.de/record/2902742

Rudman, L. A., & Glick, P. (1999). Feminized management and backlash toward agentic women: The hidden costs to women of a kinder, gentler image of middle managers. *Journal of Personality and Social Psychology*, *77*(5), 1004–1010. https://doi.org/10.1037/0022-3514.77.5.1004

Rudman, L. A., & Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues*, *57*(4), 743–762. https://doi.org/10.1111/0022-4537.00239

Rudman, L. A., Moss-Racusin, C. A., Phelan, J. E., & Nauts, S. (2012). Status incongruity and backlash effects: Defending the gender hierarchy motivates prejudice against female leaders. *Journal of Experimental Social Psychology*, *48*(1), 165–179. https://doi.org/10.1016/j.jesp.2011.10.008

Sandoval, E. B., Mubin, O., & Obaid, M. (2014). Human robot interaction and fiction: A contradiction. In M. Beetz, B. Johnston, & M.-A. Williams (Eds.), *Social Robotics* (Vol. 8755, pp. 54–63). Springer International Publishing. https://doi.org/10.1007/978-3-319-11973-1_6

Schaefer, K. (2013). *The perception and measurement of human-robot trust*. [Doctoral dissertation, University of Central Florida]. Stars Library. https://stars.library.ucf.edu/cgi/viewcontent.cgi?article=3688&context=etd

Schiffhauer, B. (2015). *Determinanten von Anthropomorphismus und ihre Bedeutung für Dehumanisierung: Zuschreibung und Absprechen von Menschlichkeit gegenüber Menschen und nicht-menschlichen Entitäten* [Determinants of anthropomorphism and their significance for dehumanization: Attribution and denial of humanity towards human and non-human entities] [Doctoral dissertation, Bielefeld University]. https://pub.uni-bielefeld.de/record/2732975

Schiffhauer, B., Bernotat, J., Eyssel, F., Bröhl, R., & Adriaans, J. (2016). Let the user decide! User preferences regarding functions, apps, and interfaces of a smart home and a service robot. In A. Agah, J.-J. Cabibihan, A. M. Howard, M. A. Salichs, & H. He (Eds.), *Social*

Robotics (Vol. 9979, pp. 971–981). Springer International Publishing. https://doi.org/10.1007/978-3-319-47437-3_95

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, *8*(4), 350–353. https://doi.org/10.1037/1040-3590.8.4.350

Schneider, W., & Chein, J. M. (2003). Controlled & automatic processing: Behavior, theory, and biological mechanisms. *Cognitive Science*, *27*(3), 525–559. https://doi.org/10.1207/s15516709cog2703_8

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*(1), 1–66. https://doi.org/10.1037/0033-295X.84.1.1

Schneider-Düker, M., & Kohler, A. (1988). Die Erfassung von Geschlechtsrollen: Ergebnisse zur deutschen Neukonstruktion des Bem Sex-Role Inventory [Assessment of sex roles: Results of a German version of the Bem Sex-Role Inventory]. *Diagnostica*, *34*(3), 256–270.

Scovel, T. (1998). *Psycholinguistics* (H. G. Widdowson, Ed.). Oxford University Press.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, *71*(2), 109–147. https://doi.org/10.1016/S0010-0277(99)00025-6

Shen, Y.-C., & Franz, E. A. (2005). Hemispheric competition in left-handers on bimanual reaction time tasks. *Journal of Motor Behavior*, *37*(1), 3–9. https://doi.org/10.3200/JMBR.37.1.3-9

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, *84*(2), 127–190. https://doi.org/10.1037/0033-295X.84.2.127

Sidanius, J., & Pratto, F. (1999). *Social dominance: An intergroup theory of social hierarchy and oppression*. Cambridge University Press.

Sidanius, J., & Pratto, F. (2012). Social dominance theory. In P. Van Lange, A. Kruglanski, & E. Higgins, *Handbook of theories of social psychology* (pp. 418–438). SAGE Publications Ltd. https://doi.org/10.4135/9781446249222.n47

Singh, R., Keshet, J., Gencaga, D., & Raj, B. (2016). The relationship of voice onset time and voice offset time to physical age. *Proceedings on International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 5390–5394. IEEE. https://doi.org/10.1109/ICASSP.2016.7472707

Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology*, *45*(4), 447–481. https://doi.org/10.1016/S0010-0285(02)00503-0

SR Research. (2009). EyeLink 1000 [Apparatus and software]. https://www.sr-research.com/

Stafford, R. Q., MacDonald, B. A., Jayawardena, C., Wegner, D. M., & Broadbent, E. (2014). Does the robot have a mind? Mind perception and attitudes towards robots predict use of an eldercare robot. *International Journal of Social Robotics*, *6*(1), 17–32. https://doi.org/10.1007/s12369-013-0186-y

Stapels, J. G., & Eyssel, F. (2021). Let's not be indifferent about robots: Neutral ratings on bipolar measures mask ambivalence in attitudes towards robots. *PLOS ONE*, *16*(1), e0244697. https://doi.org/10.1371/journal.pone.0244697

Stevens, J. P. (2009). *Applied multivariate statistics for the social sciences* (5th ed.). Routledge/Taylor & Francis Group.

Stillman, P. E., Shen, X., & Ferguson, M. J. (2018). How mouse-tracking can advance social cognitive theory. *Trends in Cognitive Sciences*, *22*(6), 531–543. https://doi.org/10.1016/j.tics.2018.03.012

Stöber, J. (1999). Die Soziale-Erwünschtheits-Skala-17 (SES-17): Entwicklung und erste Befunde zu Reliabilität und Validität [The Social Desirability Scale-17: Development and first results for reliability and validity]. *Diagnostica*, *45*(4), 173–177. https://doi.org/10.1026//0012-1924.45.4.173

Stöber, J. (2001). The Social Desirability Scale-17 (SDS-17). *European Journal of Psychological Assessment*, *17*(3), 222–232. https://doi.org/10.1027//1015-5759.17.3.222

Strait, M. K., Aguillon, C., Contreras, V., & Garcia, N. (2017). The public's perception of humanlike robots: Online social commentary reflects an appearance-based uncanny valley, a general fear of a "Technology Takeover", and the unabashed sexualization of female-gendered robots. *Proceedings of the 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (pp. 1418–1423). IEEE. https://doi.org/10.1109/ROMAN.2017.8172490

Sturt, P., Keller, F., & Dubey, A. (2010). Syntactic priming in comprehension: Parallelism effects with and without coordination. *Journal of Memory and Language*, *62*(4), 333–351. https://doi.org/10.1016/j.jml.2010.01.001

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Pearson/Allyn & Bacon.

Tamir, Y., & Nadler, A. (2007). The role of personality in social identity: Effects of field-dependence and context on reactions to threat to group distinctiveness. *Journal of Personality*, *75*(5), 927–954. https://doi.org/10.1111/j.1467-6494.2007.00461.x

Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634. https://doi.org/10.1126/science.7777863

Tay, B., Jung, Y., & Park, T. (2014). When stereotypes meet robots: The double-edge sword of robot gender and personality in human–robot interaction. *Computers in Human Behavior*, *38*, 75–84. https://doi.org/10.1016/j.chb.2014.05.014

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review*, *95*(1), 15–48.

Trueswell, J. C., Sekerina, I., Hill, N. M., & Logrip, M. L. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, *73*(2), 89–134. https://doi.org/10.1016/S0010-0277(99)00032-3

Unkelbach, C., & Fiedler, K. (2020). The challenge of diagnostic inferences from implicit measures: The case of non-evaluative influences in the evaluative priming paradigm. *Social Cognition*, *38*, 208–222. https://doi.org/10.1521/soco.2020.38.supp.s208

Vacha-Haase, T., & Thompson, B. (2004). How to estimate and interpret various effect sizes. *Journal of Counseling Psychology*, *51*(4), 473–481. https://doi.org/10.1037/0022-0167.51.4.473

van Beethoven, L. (1812, July 17). *Kunst und Musik.* BTHVN 2020 – Beethoven-Haus Bonn. https://www.beethoven.de/de/g/Kunst-und-Musik

Van Berkum, J. J. A., Holleman, B., Nieuwland, M., Otten, M., & Murre, J. (2009). Right or wrong?: The brain's fast response to morally objectionable statements. *Psychological Science*, *20*(9), 1092–1099. https://doi.org/10.1111/j.1467-9280.2009.02411.x

Van Berkum, J. J. A., Van den Brink, D., Tesink, C. M. J. Y., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, *20*(4), 580–591. https://doi.org/10.1162/jocn.2008.20054

van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. Academic Press.

Vanman, E. J., & Kappas, A. (2019). “Danger, Will Robinson!” The challenges of social robots for intergroup relations. *Social and Personality Psychology Compass*, *13*(8). https://doi.org/10.1111/spc3.12489

Velichkovsky, B. M., Dornhoefer, S. M., Pannasch, S., & Unema, P. J. A. (2000). Visual fixations and level of attentional processing. *Proceedings of the Symposium on Eye Tracking Research & Applications*, 79–85. Association for Computing Machinery (ACM). https://doi.org/10.1145/355017.355029

Velichkovsky, B. M., Rothert, A., Miniotas, D., Dornhoefer, S. M., Joos, M., & Pannasch, S. (2003). Visual fixations as a rapid indicator of hazard perception. In R. Hockey, A. W. K. Gaillard, O. Burov, & North Atlantic Treaty Organization (Eds.), *Operator functional state: The assessment and prediction of human performance degradation in complex tasks* (pp. 313–323). IOS Press.

Võ, M. L. H., Conrad, M., Kuchinke, L., Urton, K., Hofmann, M. J., & Jacobs, A. M. (2009). The Berlin Affective Word List Reloaded (BAWL-R). *Behavior Research Methods*, *41*(2), 534–538. https://doi.org/10.3758/BRM.41.2.534

Volker, M. A. (2006). Reporting effect size estimates in school psychology research. *Psychology in the Schools*, *43*(6), 653–672. https://doi.org/10.1002/pits.20176

Wang, P., Yang, Y.-P., Tan, C.-H., Chen, Q.-W., & Cantfort, T. (2017). Gender stereotype activation versus lexical semantic activation: An ERP study. *The Journal of General Psychology*, *144*(4), 283–308. https://doi.org/10.1080/00221309.2017.1310714

Waytz, A., Cacioppo, J., & Epley, N. (2010). Who sees human?: The stability and importance of individual differences in anthropomorphism. *Perspectives on Psychological Science*, *5*(3), 219–232. https://doi.org/10.1177/1745691610369336

Wegner, D. M. (1989). *Conni bears and other unwanted thoughts: Suppression, obsession, and the psychology of mental control.* Penguin Press.

Wegner, D. M., Schneider, D. J., Carter, S. R., & White, T. L. (1987). Paradoxical effects of thought suppression. *Journal of Personality and Social Psychology*, *53*(1), 5–13.

White, S. J. (2008). Eye movement control during reading: Effects of word frequency and orthographic familiarity. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(1), 205–223. https://doi.org/10.1037/0096-1523.34.1.205

Wojciszke, B. (1994). Multiple meanings of behavior: Construing actions in terms of competence or morality. *Journal of Personality and Social Psychology*, *67*(2), 222–232. https://doi.org/10.1037/0022-3514.67.2.222

Wojciszke, B., Dowhyluk, M., & Jaworski, M. (1998). Moral competence-related traits: How do they differ? *Polish Psychological Bulletin*, *29*(4), 283–294.

Wolfe, J. M. (1994). Guided search 2.0—A revised model of visual search. *Psychonomic Bulletin & Review*, *1*(2), 202–238. https://doi.org/10.3758/BF03200774

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, *5*(6), 495–501. https://doi.org/10.1038/nrn1411

Wöllstein-Leisten, A., Eisenberg, P., & Dudenredaktion (Bibliographisches Institut) (Eds.). (2016). *Duden, die Grammatik: Unentbehrlich für richtiges Deutsch* [Duden, grammar: indispensable for correct German] (9th ed.). Dudenverlag.

Wood, W., & Eagly, A. H. (2012). Biosocial construction of sex differences and similarities in behavior. In M. Zanna & J. Olson (Eds.), *Advances in experimental social psychology* (Vol. 46, pp. 55–123). Elsevier. https://doi.org/10.1016/B978-0-12-394281-4.00002-7

Wullenkord, R., & Eyssel, F. (2020a). Societal and ethical issues in HRI. *Current Robotics Reports*, *1*(3), 85–96. https://doi.org/10.1007/s43154-020-00010-9

Wullenkord, R., & Eyssel, F. (2020b). The influence of robot number on robot group perception—A call for action. *ACM Transactions on Human-Robot Interaction*, *9*(4), 1–14. https://doi.org/10.1145/3394899

Zhai, J., & Barreto, A. (2006). Stress detection in computer users based on digital signal processing of noninvasive physiological variables. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 1355–1358. IEEE. https://doi.org/10.1109/IEMBS.2006.259421

Zhang, X., Li, Q., Sun, S., & Zuo, B. (2018). The time course from gender categorization to gender-stereotype activation. *Social Neuroscience*, *13*(1), 52–60. https://doi.org/10.1080/17470919.2016.1251965

Złotowski, J., Yogeeswaran, K., & Bartneck, C. (2017). Can we control it? Autonomous robots threaten human identity, uniqueness, safety, and resources. *International Journal of Human-Computer Studies*, *100*, 48–54. https://doi.org/10.1016/j.ijhcs.2016.12.008

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*(2), 162–185. https://doi.org/10.1037/0033-2909.123.2.162

Appendix A – Pretests and Final Experimental Stimuli

Appendix A encompasses pretests and final experimental materials.

A Pretest I – Pretest Instructions

The German wording of the instructions on Pretest I (see Section 4.1) and its translation to English in a paragraph below:

> Liebe Teilnehmerin, lieber Teilnehmer, wir[39] führen eine Vorstudie durch, in der es um die Einschätzung der Geschlechtstypikalität bestimmter Begriffe geht. Hierzu präsentieren wir Ihnen eine Liste verschiedener Adjektive. Bitte schätzen Sie zunächst ein, inwiefern diese Eigenschaften in unserer Gesellschaft als typisch männlich bzw. typisch weiblich angesehen werden. Anschließend bitten wir Sie, einzuschätzen, inwiefern diese Eigenschaften in unserer Gesellschaft als positiv oder negativ angesehen werden. Bitte antworten Sie so spontan wie möglich. Es gibt keine richtigen oder falschen Antworten. Zur Bearbeitung steht Ihnen eine siebenstufige Skala zur Verfügung. Ein Pol gilt z.B. als typisch männlich, der andere als typisch weiblich. Wenn Sie denken, dass ein Adjektiv genderneutral ist, geben Sie die Mitte an. Mit den Punkten dazwischen können Sie Ihre Einschätzung abstufen. Markieren Sie bitte das Feld, das am ehesten Ihrer Einschätzung entspricht. Vielen Dank für Ihre Teilnahme!

[Dear participant, we[39] are conducting a pretest on the gender-stereotypicality of attributes. You will be shown a set of adjectives. Please estimate to what extent Western Society considers these attributes typically male or female. In the following, we ask you to estimate to what extent these attributes are considered positive or negative in Western society. Please respond as spontaneous as possible. There are no right or wrong answers. Give your responses using a 7-point Likert scale. To illustrate, one end of the spectrum indicates an attribute is evaluated as stereotypically male, the other identifies an attribute to be judged as stereotypically female. If you think an attribute is neutral in terms of gender-stereotypes or connotation, respectively, please mark the midpoint. Points in between can be used to grade your assessments. Please mark that field that best reflects your estimate. Thanks for your participation!]

---

[39] Under my supervision, students from my seminars, B.Sc.- and M.Sc.-candidates, interns, and research assistants supported me to prepare the experiments and to collect data.

A Pretest I – Results: One-Sample *t*-Tests on Attributes' Gender-Stereotypicality and Connotation

Table A1

Attributes Considered Stereotypically Male and Positively Connoted According to a One-Sample t-Test Against the Scale Midpoint of 4

Attribute	Gender-Stereotypicality						Connotation					
	M	*SD*	*t*	*df*	*p*	Cohen's *d*	*M*	*SD*	*t*	*df*	*p*	Cohen's *d*
analytisch [analytical]	3.31	1.10	-5.38	71	< .001	0.63	2.56	1.05	-11.71	71	< .001	1.38
deutlich [clear]	3.45	0.78	-5.77	66	< .001	0.71	3.32	0.95	-5.85	65	< .001	0.72
fachmännisch [expert]	2.64	1.00	-11.17	66	< .001	1.37	2.08	0.92	-16.78	64	< .001	2.08
fähig [capable]	3.75	0.87	-2.44	71	.017	0.29	1.86	1.06	-17.02	70	< .001	2.02
gekonnt [skillful]	3.70	0.95	-2.63	70	.010	0.31	2.08	1.31	-12.43	71	< .001	1.47
gelassen [calm]	3.38	1.03	-5.16	71	< .001	0.61	2.66	1.11	-10.18	70	< .001	1.21
heldenhaft [heroic]	2.40	1.24	-10.51	66	< .001	1.28	1.85	0.90	-19.45	65	< .001	2.39
lässig [casual]	3.00	0.90	-9.39	71	< .001	1.11	3.40	0.95	-5.12	66	< .001	0.63
logisch [logical]	2.83	0.95	-10.43	71	< .001	1.23	2.45	0.96	-13.37	68	< .001	1.61
lösungsorientiert [solution-oriented]	3.49	1.11	-3.76	66	< .001	0.46	2.08	0.81	-19.31	65	< .001	2.38
mutig [brave]	2.64	1.11	-10.01	66	< .001	1.22	1.86	0.92	-18.54	63	< .001	2.32
pragmatisch [pragmatic]	2.83	1.00	-9.47	65	< .001	1.17	2.79	1.02	-9.70	65	< .001	1.19
praktisch [practical]	2.88	1.39	-6.60	66	< .001	0.81	2.78	1.10	-8.93	64	< .001	1.11
professionell [professional]	3.65	1.01	-2.92	71	.005	0.34	1.68	0.98	-20.16	71	< .001	2.38
rational [rational]	2.90	1.27	-7.34	71	< .001	0.87	2.74	1.10	-9.75	71	< .001	1.15
risikobereit [risk-taking]	2.42	0.90	-14.30	65	< .001	1.76	3.70	1.04	-2.37	65	.021	0.29
sachlich [objective]	3.28	0.94	-6.53	71	< .001	0.77	2.86	1.16	-8.27	70	< .001	0.98
schnell [quick]	3.67	0.79	-3.42	65	.001	0.42	3.02	0.91	-8.72	64	< .001	1.08
selbstbewusst [confident]	3.39	0.83	-6.01	66	< .001	0.73	1.65	0.69	-27.33	64	< .001	3.39
sicher [confident]	3.37	0.83	-6.17	66	< .001	0.75	2.17	0.87	-17.13	65	< .001	2.11
spontan [spontaneous]	3.70	0.98	-2.52	65	.014	0.31	2.71	1.12	-9.34	65	< .001	1.15
stark [strong]	2.42	1.28	-10.36	70	< .001	1.23	2.13	1.10	-14.47	71	< .001	1.71

Note. Gender-Stereotypicality: 1 = stereotypically male, 7 = stereotypically female; Connotation: 1 = positive, 7 = negative. The attribute's translation to English is inserted in parentheses below.[40]

(continued)

[40] In tables containing means on gender-stereotypicality and connotation, the items were sorted alphabetically. In tables containing only means on gender-stereotypicality, means were sorted in ascending order.

Table A1

Attributes Considered Stereotypically Male and Positively Connoted According to a One-Sample t-Test Against the Scale Midpoint of 4 (continued)

Attribute	Gender-Stereotypicality						Connotation					
	M	*SD*	*t*	*df*	*p*	Cohen's *d*	*M*	*SD*	*t*	*df*	*p*	Cohen's *d*
stolz [proud]	3.30	1.28	-4.63	70	< .001	0.55	3.48	1.39	-3.12	68	.003	0,38
strategisch [strategic]	3.10	0.84	-8.75	66	< .001	1.07	2.95	1.01	-8.38	64	< .001	1,04
überzeugt [convinced]	3.47	0.95	-4.72	71	< .001	0.56	2.60	1.17	-10.17	71	< .001	1,20
unerschrocken [fearless]	2.90	0.95	-9.06	61	< .001	1.15	2.85	1.05	-8.57	60	< .001	1,10

Note. Gender-Stereotypicality: 1 = stereotypically male, 7 = stereotypically female; Connotation: 1 = positive, 7 = negative. The attribute's translation to English is inserted in parentheses below.

Table A2

Attributes Considered Stereotypically Male and Negatively Connoted According to a One-Sample t-Test Against the Scale Midpoint of 4

Attribute	Gender-Stereotypicality						Connotation					
	M	*SD*	*t*	*df*	*p*	Cohen's *d*	*M*	*SD*	*t*	*df*	*p*	Cohen's *d*
achtlos [careless]	3.47	0.90	-4.96	71	< .001	0.59	6.24	0.80	23.84	71	< .001	2.81
aggressiv [aggressive]	2.56	1.07	-11.42	71	< .001	1.35	6.33	1.06	18.65	71	< .001	2.20
angeberisch [swanky]	2.51	0.91	-13.41	66	< .001	1.64	6.25	0.79	22.89	64	< .001	2.84
animalisch [animal]	2.78	1.39	-7.48	71	< .001	0.88	5.00	1.40	6.04	71	< .001	0.71
arrogant [arrogant]	3.72	1.01	-2.29	66	.025	0.28	6.29	0.89	20.88	65	< .001	2.57
bedenkenlos [regardless]	3.34	0.76	-7.06	64	< .001	0.88	5.02	1.05	7.77	64	< .001	0.96
cholerisch [irascible]	2.72	1.35	-7.80	66	< .001	0.95	6.18	1.20	14.71	64	< .001	1.83
forsch [brisk]	3.26	1.34	-4.65	71	< .001	0.55	4.67	1.62	3.50	71	.001	0.41
gedankenlos [thoughtless]	3.75	0.89	-2.32	66	.023	0.28	5.30	0.82	12.88	65	< .001	1.59
gefühllos [unemotional]	3.06	0.92	-8.37	66	< .001	1.02	6.31	0.79	23.58	64	< .001	2.93
geringschätzig [disparaging]	3.79	0.71	-2.42	66	.018	0.30	5.95	1.14	13.90	65	< .001	1.71
gierig [greedy]	3.75	0.80	-2.58	66	.012	0.32	6.20	0.73	24.19	64	< .001	3.00
grob [rude]	2.70	0.83	-13.08	70	< .001	1.55	5.72	1.05	13.90	71	< .001	1.64
grobmotorisch [gross]	2.78	1.19	-8.41	66	< .001	1.03	5.28	0.88	11.76	64	< .001	1.46
harsch [harsh]	3.22	1.02	-6.45	71	< .001	0.76	5.46	1.18	10.37	69	< .001	1.24

Note. Gender-Stereotypicality: 1 = stereotypically male, 7 = stereotypically female; Connotation: 1 = positive, 7 = negative. The attribute's translation to English is inserted in parentheses below.

(continued)

Table A2

Attributes Considered Stereotypically Male and Negatively Connoted According to a One-Sample t-Test Against the Scale Midpoint of 4 (continued)

Attribute	Gender-Stereotypicality						Connotation					
	M	*SD*	*t*	*df*	*p*	Cohen's *d*	*M*	*SD*	*t*	*df*	*p*	Cohen's *d*
herablassend [condescending]	3.63	1.14	-2.78	71	.007	0.33	6.22	0.98	18.73	68	< .001	2.26
lautstark [loud]	2.69	0.86	-12.19	64	< .001	1.51	5.06	0.97	8.84	65	< .001	1.09
leichtsinnig [reckless]	2.71	1.26	-8.69	71	< .001	1.02	5.52	1.03	12.49	70	< .001	1.48
machtbesessen [power-obsessed]	2.53	1.19	-10.53	71	< .001	1.24	6.00	1.15	14.70	70	< .001	1.75
missmutig [morose]	3.77	0.67	-2.74	65	.008	0.34	5.62	0.99	13.32	65	< .001	1.64
prahlerisch [boastful]	2.65	1.19	-9.62	71	< .001	1.13	6.08	0.99	17.87	71	< .001	2.11
respektlos [disrespectful]	3.54	0.64	-5.96	66	< .001	0.73	6.36	1.24	15.54	65	< .001	1.91
rücksichtslos [ruthless]	3.15	0.89	-7.81	66	< .001	0.95	6.36	0.91	21.21	65	< .001	2.61
rüpelhaft [loutish]	2.25	0.88	-16.31	66	< .001	1.99	6.25	0.73	24.82	64	< .001	3.08
schroff [rough]	2.96	0.93	-9.21	66	< .001	1.13	5.94	0.96	16.43	65	< .001	2.02
überheblich [pretentious]	3.38	0.96	-5.55	71	< .001	0.65	6.10	1.06	16.73	71	< .001	1.97
unachtsam [heedless]	3.41	1.04	-4.81	70	< .001	0.57	5.93	1.01	16.19	71	< .001	1.91
ungehalten [indignant]	3.33	0.96	-5.87	71	< .001	0.69	5.67	1.30	10.88	71	< .001	1.28
unhöflich [impolite]	3.72	0.57	-4.06	66	< .001	0.50	6.29	0.70	26.37	64	< .001	3.27
unordentlich [untidy]	3.03	1.21	-6.79	70	< .001	0.81	5.79	1.17	12.95	71	< .001	1.53
unverschämt [impertinent]	3.13	1.19	-6.26	71	< .001	0.74	6.49	0.87	24.19	71	< .001	2.85
unvorsichtig [incautious]	3.24	0.92	-6.75	66	< .001	0.83	5.62	0.94	14.00	65	< .001	1.72
unzuverlässig [unreliable]	3.43	0.96	-5.02	71	< .001	0.59	6.59	0.93	23.36	70	< .001	2.77
verantwortungslos [irresponsible]	3.47	0.75	-5.75	65	< .001	0.71	6.47	0.95	21.16	65	< .001	2.61
verständnislos [uncomprehending]	3.51	1.20	-3.44	71	.001	0.41	5.76	1.28	11.66	71	< .001	1.37

Note. Gender-Stereotypicality: 1 = stereotypically male, 7 = stereotypically female; Connotation: 1 = positive, 7 = negative. The attribute's translation to English is inserted in parentheses below.

Table A3

Attribute Considered Stereotypically Male and Neither Positively Nor Negatively Connoted According to a One-Sample t-Test Against the Scale Midpoint of 4

Attribute	Gender-Stereotypicality						Connotation					
	M	*SD*	*t*	*df*	*p*	Cohen's *d*	*M*	*SD*	*t*	*df*	*p*	Cohen's *d*
nachdrücklich [insistent]	3.55	0.95	-3.98	70	<.001	0.47	3.88	1.01	-1.05	64	.295	0.12

Note. Gender-Stereotypicality: 1 = stereotypically male, 7 = stereotypically female; Connotation: 1 = positive, 7 = negative. The attribute's translation to English is inserted in parentheses below.

Table A4

Attributes Considered Stereotypically Female and Positively Connoted According to a One-Sample t-Test Against the Scale Midpoint of 4

Attribute	Gender-Stereotypicality						Connotation					
	M	*SD*	*t*	*df*	*p*	Cohen's *d*	*M*	*SD*	*t*	*df*	*p*	Cohen's *d*
achtsam [heedful]	5.03	0.85	9.89	66	< .001	1.21	2.91	1.05	-8.45	65	< .001	1.04
aufmerksam [attentive]	4.60	1.12	4.52	71	< .001	0.53	1.64	0.84	-23.73	71	< .001	2.80
bedächtig [thoroughful]	4.56	0.87	5.43	70	< .001	0.64	3.12	1.12	-6.57	68	< .001	0.79
ehrfürchtig [reverent]	4.31	0.87	2.94	66	.005	0.36	3.52	1.04	-3.78	65	< .001	0.47
einfühlsam [empathetic]	5.42	0.99	11.65	65	< .001	1.43	2.22	0.91	-15.81	64	< .001	1.96
elegant [elegant]	5.32	0.93	11.50	65	< .001	1.42	2.38	1.19	-11.10	65	< .001	1.37
emotional [emotional]	5.89	0.88	18.19	71	< .001	2.14	3.57	1.22	-2.99	71	.004	0.35
euphorisch [euphoric]	4.69	1.00	5.88	71	< .001	0.69	3.19	1.37	-4.99	71	< .001	0.59
fasziniert [fascinated]	4.39	0.87	3.66	65	.001	0.45	2.79	1.17	-8.41	65	< .001	1.04
feinfühlig [sensitively]	5.67	1.09	13.00	71	< .001	1.53	2.21	1.20	-12.69	71	< .001	1.50
fingerfertig [dexterous]	4.54	1.09	4.03	66	< .001	0.49	2.64	1.06	-10.43	65	< .001	1.28
fleißig [diligent]	4.68	1.20	4.83	71	< .001	0.57	1.54	0.79	-26.54	71	< .001	3.13
freundlich [friendly]	4.40	0.74	4.46	66	< .001	0.55	1.57	0.90	-21.75	64	< .001	2.70
fürsorglich [caring]	5.51	1.05	12.26	71	< .001	1.45	1.93	1.14	-15.33	70	< .001	1.82
geduldig [patient]	4.97	1.29	6.40	71	< .001	0.75	1.99	1.19	-14.33	71	< .001	1.69
gefühlvoll [emotional]	5.60	1.25	10.82	71	< .001	1.28	2.48	1.13	-11.32	70	< .001	1.34
geheimnisvoll [secretive]	4.51	0.88	4.74	66	< .001	0.58	3.63	0.89	-3.33	64	.001	0.41

Note. Gender-Stereotypicality: 1 = stereotypically male, 7 = stereotypically female; Connotation: 1 = positive, 7 = negative. The attribute's translation to English is inserted in parentheses below.

(continued)

Table A4

Attributes Considered Stereotypically Female and Positively Connoted According to a One-Sample t-Test Against the Scale Midpoint of 4 (continued)

Attribute	Gender-Stereotypicality						Connotation					
	M	*SD*	*t*	*df*	*p*	Cohen's *d*	*M*	*SD*	*t*	*df*	*p*	Cohen's *d*
genau [accurate]	4.21	0.84	2.03	66	.047	0.25	2.65	1.00	-10.96	65	< .001	1.35
gewissenhaft [conscientious]	4.46	1.02	3.81	71	< .001	0.45	1.60	0.83	-24.46	71	< .001	2.88
gründlich [thorough]	4.82	1.00	6.73	66	< .001	0.82	2.61	1.04	-10.93	65	< .001	1.35
herzlich [cordial]	4.94	1.07	7.16	65	< .001	0.88	2.00	0.94	-17.20	65	< .001	2.12
hilfsbereit [helpful]	4.83	1.10	6.42	71	< .001	0.76	1.39	0.85	-25.73	70	< .001	3.05
hingebungsvoll [devoted]	5.22	1.00	10.05	66	< .001	1.23	2.38	0.95	-13.75	64	< .001	1.71
höflich [polite]	4.30	0.78	3.14	65	.003	0.39	1.89	1.11	-15.40	65	< .001	1.90
kokett [coquettish]	5.19	1.05	9.33	66	< .001	1.14	3.58	1.19	-2.90	65	.005	0.36
kreativ [creative]	5.01	1.07	8.06	71	< .001	0.95	1.93	1.14	-15.37	71	< .001	1.81
kunstfertig [skillful]	4.83	1.22	5.79	71	< .001	0.68	2.59	1.26	-9.42	70	< .001	1.12
leidenschaftlich [passionate]	4.49	0.96	4.20	66	< .001	0.51	2.38	1.10	-11.84	64	< .001	1.47
liebevoll [loving]	5.04	0.96	8.91	66	< .001	1.09	1.77	0.97	-18.59	65	< .001	2.29
modebewusst [fashionable]	5.58	1.03	13.03	71	< .001	1.54	3.06	1.31	-6.08	70	< .001	0.72
nachdenklich [pensive]	4.37	0.93	3.27	66	.002	0.40	3.75	0.97	-2.05	64	.045	0.25
neugierig [curious]	4.24	0.66	2.99	65	.004	0.37	2.80	1.14	-8.53	65	< .001	1.05
ordentlich [proper]	4.94	1.09	7.38	71	< .001	0.87	2.09	0.92	-17.28	68	< .001	2.08
ordnungsgemäß [proper]	4.32	0.98	2.64	65	.010	0.33	2.65	1.10	-9.94	65	< .001	1.22
raffiniert [cunning]	4.28	0.83	2.79	66	.007	0.34	2.79	1.27	-7.75	65	< .001	0.95
respektvoll [respectful]	4.28	1.05	2.24	71	.028	0.26	1.51	1.03	-20.39	71	< .001	2.40
rücksichtsvoll [considerate]	5.15	0.99	9.90	71	< .001	1.17	1.83	1.14	-15.88	68	< .001	1.91
sanft [soft]	5.51	0.93	13.31	66	< .001	1.63	2.43	0.97	-13.07	64	< .001	1.62
sauber [clean]	5.17	1.10	8.99	71	< .001	1.06	2.00	0.99	-17.10	70	< .001	2.03
sensibel [sensitive]	5.39	0.88	12.48	61	< .001	1.59	3.50	1.21	-3.19	59	.002	0.41
sorgfältig [careful]	4.67	0.86	6.40	66	< .001	0.78	2.35	1.03	-13.02	65	< .001	1.60
sorgsam [deligent]	5.26	0.99	10.80	71	< .001	1.27	2.31	1.27	-11.28	71	< .001	1.33
stilvoll [stylish]	4.63	0.88	5.80	66	< .001	0.71	2.38	0.86	-15.14	64	< .001	1.88

Note. Gender-Stereotypicality: 1 = stereotypically male, 7 = stereotypically female; Connotation: 1 = positive, 7 = negative. The attribute's translation to English is inserted in parentheses below.

(continued)

Table A4

Attributes Considered Stereotypically Female and Positively Connoted According to a One-Sample t-Test Against the Scale Midpoint of 4 (continued)

Attribute	Gender-Stereotypicality						Connotation					
	M	*SD*	*t*	*df*	*p*	Cohen's *d*	*M*	*SD*	*t*	*df*	*p*	Cohen's *d*
umsichtig [prudent]	4.56	0.86	5.29	65	< .001	0.65	2.38	0.84	-15.47	64	< .001	1.92
verantwortungsvoll [responsible]	4.57	1.17	4.12	71	< .001	0.49	1.35	0.68	-32.29	68	< .001	3.89
verständnisvoll [sympathetic]	5.17	0.89	10.68	65	< .001	1.32	2.00	0.79	-20.40	64	< .001	2.53
vorausschauend [foresighted]	4.33	1.22	2.31	71	.024	0.27	2.28	0.96	-15.10	70	< .001	1.79
vorsorglich [precautionary]	4.73	1.06	5.85	70	<.001	0.69	2.42	0.99	-13.58	71	< .001	1.60
wohlüberlegt [deliberate]	4.42	0.89	3.84	66	< .001	0.47	2.33	0.95	-14.25	65	< .001	1.75
wortgewandt [eloquent]	4.59	1.37	3.64	70	.001	0.43	2.11	1.22	-13.17	71	< .001	1.55
zärtlich [tender]	5.40	1.05	10.99	66	< .001	1.34	2.33	1.17	-11.59	65	< .001	1.43
zuverlässig [reliable]	4.76	1.11	5.86	71	< .001	0.69	1.39	0.79	-27.43	68	< .001	3.30
zuvorkommend [obliging]	4.33	1.04	2.60	66	.012	0.32	1.95	0.80	-20.65	64	< .001	2.56

Note. Gender-Stereotypicality: 1 = stereotypically male, 7 = stereotypically female; Connotation: 1 = positive, 7 = negative. The attribute's translation to English is inserted in parentheses below.

Table A5

Attributes Considered Stereotypically Female and Negatively Connoted According to a One-Sample t-Test Against the Scale Midpoint of 4

Attribute	Gender-Stereotypicality						Connotation					
	M	*SD*	*t*	*df*	*p*	Cohen's *d*	*M*	*SD*	*t*	*df*	*p*	Cohen's *d*
ängstlich [anxious]	5.42	1.15	10.38	70	< .001	1.23	5.36	1.08	10.54	69	< .001	1.26
eingebildet [conceited]	4.65	1.14	4.81	70	< .001	0.57	6.32	0.99	19.87	71	< .001	2.34
heimtückisch [insidious]	4.54	1.03	4.21	64	< .001	0.52	6.62	0.58	36.48	64	< .001	4.53
hektisch [hectic]	4.76	0.91	7.11	71	< .001	0.84	5.37	1.23	9.33	70	< .001	1.11
hinterlistig [cunning]	4.93	1.19	6.63	71	< .001	0.78	6.49	1.00	21.04	70	< .001	2.50
intrigant [scheming]	5.15	0.91	10.35	66	< .001	1.26	6.15	1.06	16.55	65	< .001	2.04
naiv [naïve]	5.21	0.91	10.83	66	< .001	1.32	5.63	0.74	17.74	64	< .001	2.20
nervös [nervous]	4.79	0.99	6.77	71	< .001	0.80	5.35	1.13	10.14	71	< .001	1.20
orientierungslos [disoriented]	4.47	1.21	3.31	71	.001	0.39	5.54	1.07	12.18	71	< .001	1.44
panisch [panic]	5.15	0.93	10.17	66	< .001	1.24	5.94	0.77	20.35	64	< .001	2.52

Note. Gender-Stereotypicality: 1 = stereotypically male, 7 = stereotypically female; Connotation: 1 = positive, 7 = negative. The attribute's translation to English is inserted in parentheses below.

(continued)

Table A5

Attributes Considered Stereotypically Female and Negatively Connoted According to a One-Sample t-Test Against the Scale Midpoint of 4 (continued)

Attribute	Gender-Stereotypicality						Connotation					
	M	*SD*	*t*	*df*	*p*	Cohen's *d*	*M*	*SD*	*t*	*df*	*p*	Cohen's *d*
penibel [meticulously]	5.08	1.13	8.08	70	< .001	0.96	4.42	1.31	2.70	71	.009	0.32
schüchtern [shy]	4.85	1.13	6.34	71	.001	0.75	4.68	0.98	5.67	67	< .001	0.69
selbstzweifelnd [self-doubting]	4.92	1.18	6.57	71	< .001	0.77	5.31	1.29	8.62	71	< .001	1.02
überempfindlich [hypersensitive]	5.49	0.89	13.67	66	< .001	1.67	5.74	0.71	19.98	65	< .001	2.46
übervorsichtig [overcautious]	5.59	0.93	13.93	65	< .001	1.72	5.30	0.81	12.80	63	< .001	1.60
unfähig [inapt]	4.18	0.64	2.41	71	.018	0.28	6.01	1.07	16.00	71	< .001	1.89
ungeschickt [clumsy]	4.48	0.70	5.55	66	< .001	0.68	5.59	0.82	15.72	65	< .001	1.94
unsicher [unconfident]	4.88	0.75	9.62	66	< .001	1.18	5.29	0.79	13.27	64	< .001	1.65
verlogen [mendacious]	4.49	0.91	4.43	66	< .001	0.54	6.58	0.79	26.43	64	< .001	3.28
verträumt [dreamy]	4.78	1.26	5.24	71	< .001	0.62	4.42	1.13	3.15	70	.002	0.37
zurückhaltend [unobstrusive]	4.76	0.72	8.32	61	< .001	1.06	4.70	0.78	7.04	60	< .001	0.90

Note. Gender-Stereotypicality: 1 = stereotypically male, 7 = stereotypically female; Connotation: 1 = positive, 7 = negative. The attribute's translation to English is inserted in parentheses below.

Table A6

Attributes Considered Stereotypically Female and Neither Positively Nor Negatively Connoted According to a One-Sample t-Test Against the Scale Midpoint of 4

Attribute	Gender-Stereotypicality						Connotation					
	M	*SD*	*t*	*df*	*p*	Cohen's *d*	*M*	*SD*	*t*	*df*	*p*	Cohen's *d*
besorgt [worried]	5.34	0.96	11.43	66	< .001	1.40	4.12	0.97	1.02	65	.313	0.13
leise [quiet]	4.76	1.09	5.92	71	< .001	0.70	4.17	1.08	1.32	70	.192	0.16
skeptisch [sceptical]	4.29	1.18	2.10	71	< .001	0.25	4.24	1.16	1.73	71	.088	0.20
vorsichtig [cautious]	5.10	0.86	10.57	66	< .001	1.29	3.80	0.85	-1.89	65	.063	0.23

Note. Gender-Stereotypicality: 1 = stereotypically male, 7 = stereotypically female; Connotation: 1 = positive, 7 = negative. The attribute's translation to English is inserted in parentheses below.

Table A7

Attributes Considered Stereotypically Gender-Neutral and Positively Connoted According to a One-Sample t-Test Against the Scale Midpoint of 4

Attribute	Gender-Stereotypicality						Connotation					
	M	*SD*	*t*	*df*	*p*	Cohen's *d*	*M*	*SD*	*t*	*df*	*p*	Cohen's *d*
akribisch [meticulous]	4.15	0.86	1.43	65	.159	0.18	3.39	1.23	-4.01	65	< .001	0.49
durchdacht [sophisticated]	3.72	1.31	-1.79	71	.077	0.21	2.54	1.37	-9.01	71	< .001	1.06
ehrlich [honest]	4.06	0.95	0.50	71	.620	0.06	1.58	1.07	-19.14	71	< .001	2.26
flink [swift]	4.06	0.80	0.61	66	.541	0.08	2.92	0.85	-10.17	64	< .001	1.26
geschickt [skillfull]	3.94	0.95	-0.51	66	.609	0.06	2.14	0.84	-18.04	65	< .001	2.22
großzügig [generous]	3.86	1.05	-1.13	70	.260	0.13	2.07	1.18	-13.57	68	< .001	1.63
inspirierend [inspiring]	4.02	0.98	0.13	65	.901	0.02	3.61	1.31	-2.44	65	.017	0.30
klug [clever]	4.07	0.82	0.73	70	.470	0.09	1.56	0.96	-21.55	71	< .001	2.54
kompetent [competent]	3.85	0.83	-1.56	71	.124	0.18	1.40	0.76	-28.89	71	< .001	3.41
konzentriert [focused]	4.07	0.61	1.00	66	.321	0.12	2.20	1.00	-14.71	65	< .001	1.81
kritisch [critical]	4.05	0.95	0.39	65	.699	0.05	3.65	1.23	-2.29	65	.025	0.28
motiviert [motivated]	4.06	0.68	0.73	65	.469	0.09	1.66	0.71	-26.43	64	< .001	3.28
musikalisch [musical]	4.17	0.90	1.56	71	.122	0.18	2.35	1.06	-13.19	71	< .001	1.56
pünktlich [punctual]	4.08	1.16	0.61	71	.544	0.07	1.57	0.92	-22.50	71	< .001	2.65
routiniert [experienced]	4.26	1.10	1.96	69	.054	0.23	2.97	1.22	-7.14	71	< .001	0.84
virtuos [virtuoso]	4.00	0.72	< 0.01	65	1.000	< 0.01	3.18	1.36	-4.90	65	< .001	0.60
zielstrebig [determined]	3.86	1.14	-1.03	71	.306	0.12	1.68	0.83	-23.17	68	< .001	2.79
zukunftsorientiert [future-oriented]	4.06	0.92	0.53	66	.597	0.07	2.21	0.85	-17.08	65	< .001	2.10

Note. Gender-Stereotypicality: 1 = stereotypically male, 7 = stereotypically female; Connotation: 1 = positive, 7 = negative. The attribute's translation to English is inserted in parentheses below.

Table A8

Attributes Considered Stereotypically Gender-Neutral and Negatively Connoted According to a One-Sample t-Test Against the Scale Midpoint of 4

Attribute	Gender-Stereotypicality						Connotation					
	M	*SD*	*t*	*df*	*p*	Cohen's *d*	*M*	*SD*	*t*	*df*	*p*	Cohen's *d*
genervt [annoyed]	4.18	1.14	1.34	71	.184	0.16	6.22	0.92	20.44	71	< .001	2.41
inkompetent [incompetent]	4.13	0.67	1.58	71	.118	0.19	6.08	1.28	13.67	70	< .001	1.62
langsam [slow]	3.94	0.67	-0.73	66	.469	0.09	5.52	0.99	12.45	64	< .001	1.54
oberflächlich [superficial]	4.13	1.31	0.84	66	.405	0.10	5.86	1.02	14.83	65	< .001	1.83
spöttisch [mocking]	4.16	1.02	1.31	66	.194	0.16	6.05	0.81	20.46	65	< .001	2.52
tollpatschig [clumsy]	4.17	1.33	1.06	71	.292	0.13	5.03	1.17	7.40	70	< .001	0.88
unbeholfen [inapt]	4.16	1.08	1.24	66	.218	0.15	5.42	0.77	15.11	65	< .001	1.86
ungeduldig [impatient]	3.82	0.94	-1.57	65	.122	0.19	5.58	0.86	14.88	65	< .001	1.83
ungläubig [incredulous]	3.88	0.75	-1.42	71	.161	0.17	4.77	1.20	5.31	68	< .001	0.64

Note. Gender-Stereotypicality: 1 = stereotypically male, 7 = stereotypically female; Connotation: 1 = positive, 7 = negative. The attribute's translation to English is inserted in parentheses below.

Table A9

Attribute Considered Stereotypically Gender-Neutral and Neither Positively Nor Negatively Connoted According to a One-Sample t-Test Against the Scale Midpoint of 4

Attribute	Gender-Stereotypicality						Connotation					
	M	*SD*	*t*	*df*	*p*	Cohen's *d*	*M*	*SD*	*t*	*df*	*p*	Cohen's *d*
prüfend [checking]	4.15	1.08	1.13	66	.261	0.14	3.77	0.97	-1.90	65	.062	0.23

Note. Gender-Stereotypicality: 1 = stereotypically male, 7 = stereotypically female; Connotation: 1 = positive, 7 = negative. The attribute's translation to English is inserted in parentheses below.

A Pretest II – Pretest Instructions

The German wording instructions on Pretest II (see Section 4.2) and its translation to English in a paragraph below:

> Liebe Teilnehmerin, lieber Teilnehmer, wir führen eine Vorstudie durch, in der es um die Einschätzung der Geschlechtstypikalität bestimmter Begriffe geht. Hierzu präsentieren wir Ihnen einzelne Satzbausteine. Bitte schätzen Sie ein, inwiefern diese Satzbausteine in unserer Gesellschaft als typisch männlich, typisch weiblich oder gender-neutral angesehen werden. Berücksichtigen Sie dabei nicht, welche Satzbausteine Ihnen vorher gezeigt wurden, sondern bewerten Sie jeden Satzbaustein für sich, auch wenn diese Ihnen als zusammenpassend erscheinen. So wird z. B. das Nomen „das Lineal" einzeln bewertet und das nachfolgende mögliche Verb „wird genutzt" wird ebenfalls unabhängig vom vorangegangenen Satzbaustein bewertet. Bitte antworten Sie so spontan wie möglich. Es gibt keine richtigen oder falschen Antworten. Zur Bearbeitung steht Ihnen eine siebenstufige Skala zur Verfügung. Ein Pol gilt als typisch männlich (= 1), der andere als typisch weiblich (= 7). Wenn Sie denken, dass ein Satzbaustein gender-neutral ist, geben Sie die Mitte an (= 4). Mit den Punkten dazwischen können Sie Ihre Einschätzung abstufen. Markieren Sie bitte das Feld, das am ehesten Ihrer Einschätzung entspricht. Die Bearbeitung dieses Fragebogens wird ca. 20 Minuten dauern. Wenn benötigt, erhalten Sie dafür eine halbe Versuchspersonenstunde. Geben Sie dazu bitte am Ende der Studie Ihre Versuchspersonennummer an. Falls Sie keine Versuchspersonenstunden benötigen, lassen Sie dieses Feld einfach frei. Vielen Dank für Ihre Teilnahme!

[Dear participant, we conduct a pilot study to investigate to what extent a list of verbs and nouns is considered gender-stereotypical. To do so, you will be presented single words of a sentence. Please estimate to what extent Western Society considers these words stereotypically male, stereotypically female, or stereotypically gender-neutral. While evaluating a word's gender-stereotypicality, please neglect the word you might have read antecedently. Evaluate each word independently even if it seemingly fits to its antecedent. To illustrate, the noun "the ruler" and the verb that might follow, e.g., "is used" are both to be evaluated independently.

Please respond as spontaneous as possible. There are no right or wrong answers. Please use a 7-point bipolar Likert scale to give your responses. One pole of the scale marks a word as stereotypically male (= 1) the other pole marks a word being judged as stereotypically female (= 7). Please mark the midpoint of the scale if you think a word is considered neutral in terms of gender stereotypes (= 4). The points in-between can be used to grade your estimates. Please mark the field which best represents your estimates. It might take about 20 minutes to complete the survey. If needed, you can get 0.5 course credits for participation. Please indicate your ID at the end of the survey if you need course credits. If no course credits are needed, leave this field blank. Thanks for your participation!]

A Pretest II – Results: One-Sample *t*-Tests on Noun and Verb Phrases' Gender-Stereotypicality

Table A10

Noun Phrases Considered Stereotypically Gender-Neutral According to a One-Sample t-Test Against the Scale Midpoint of 4

Noun Phrase	Gender-Stereotypicality					
	M	*SD*	*t*	*df*	*p*	Cohen's *d*
das Fachbuch [the reference book]	3.66	1.10	-1.92	37	.062	0.31
das Brettspiel [the board game]	3.70	0.95	-1.83	32	.077	0.32
das Dokument [the document]	3.73	0.94	-1.66	32	.107	0.29
das Brot [the bread]	3.74	0.83	-1.96	37	.058	0.32
das Haus [the house]	3.74	0.86	-1.89	37	.067	0.31
das Opossum [the opossum]	3.79	0.70	-1.75	32	.090	0.31
das Regal [the shelf]	3.79	0.96	-1.35	37	.186	0.22
das Dreieck [the triangle]	3.82	0.53	-1.98	32	.056	0.35
das Labyrinth [the labyrinth]	3.82	0.58	-1.79	32	.083	0.31
das Getränk [the drink]	3.82	0.61	-1.87	37	.070	0.30
das Ladegerät [the charger]	3.82	0.65	-1.74	37	.090	0.28
das Naturereignis [the natural event]	3.82	0.68	-1.53	32	.136	0.27
das Trapez [the trapezoid]	3.82	0.88	-1.18	32	.245	0.21
das Spielzeug [the toy]	3.82	1.10	-0.95	32	.351	0.17
das Radio [the radio]	3.84	0.72	-1.36	37	.183	0.22
das Viereck [the quadrilateral]	3.85	0.51	-1.72	32	.096	0.30
das Futter [the feed]	3.85	0.67	-1.31	32	.201	0.23
das Ei [the egg]	3.87	1.12	-0.73	37	.473	0.12
das Handy [the cellphone]	3.88	0.49	-1.44	32	.160	0.25
das Glücksrad [the prize wheel]	3.88	0.82	-0.85	32	.402	0.15
das Aquarium [the aquarium]	3.91	0.95	-0.55	32	.585	0.10
das Lineal [the ruler]	3.92	0.43	-1.14	37	.262	0.19
das Radiergummi [the eraser]	3.92	0.67	-0.72	37	.474	0.12

Note. Gender-Stereotypicality: 1 = stereotypically male, 7 = stereotypically female. The noun phrase's translation to English is inserted in parentheses below.[40]

(continued)

Table A10

Noun Phrases Considered Stereotypically Gender-Neutral According to a One-Sample t-Test Against the Scale Midpoint of 4 (continued)

Noun Phrase	Gender-Stereotypicality					
	M	*SD*	*t*	*df*	*p*	Cohen's *d*
das Publikum [the audience]	3.94	0.24	-1.44	32	.160	0.25
das Zimmer [the room]	3.95	0.23	-1.43	37	.160	0.23
das Mobiliar [the furniture]	3.97	0.82	-0.20	37	.845	0.03
das Wasser[41] [the water]	4.00	< 0.01	-	-	-	
das Glas [the glass]	4.00	0.62	< 0.01	37	1.000	< 0.01
das Tier [the animal]	4.00	0.74	< 0.01	37	1.000	< 0.01
das Album [the album]	4.00	0.87	< 0.01	32	1.000	< 0.01
das Blatt [the sheet]	4.05	0.23	1.43	37	.160	0.23
das Portemonnaie [the wallet]	4.05	0.73	0.44	37	.661	0.07
das Fenster [the window]	4.05	0.77	0.42	37	.676	0.07
das Paket [the parcel]	4.05	0.93	0.35	37	.729	0.06
das Licht [the light]	4.06	0.70	0.49	32	.625	0.09
das Apartment [the apartment]	4.06	1.00	0.35	32	.730	0.06
das Heft [the booklet]	4.09	0.52	1.00	32	.325	0.17
das Klavier [the piano]	4.09	0.98	0.53	32	.598	0.09
das Brötchen [the roll]	4.11	0.80	0.81	37	.422	0.13
das Handtuch [the towel]	4.15	0.71	1.22	32	.231	0.21
das Eichhörnchen [the squirrel]	4.15	0.80	1.09	32	.282	0.19
das Fischfilet [the fish fillet]	4.15	0.97	0.90	32	.377	0.16
das Xylophon [the xylophone]	4.16	0.50	1.97	37	.057	0.32
das Museum [the museum]	4.18	0.58	1.79	32	.083	0.31
das Sofa [the sofa]	4.18	0.58	1.79	32	.083	0.31
das Toilettenpapier [the toilet paper]	4.18	0.58	1.79	32	.083	0.31
das Taschentuch [the handkerchief]	4.18	0.73	1.44	32	.160	0.25
das Plakat [the poster]	4.18	0.77	1.48	37	.147	0.24
das Tablet [the tablet]	4.26	0.86	1.89	37	.067	0.31
das Gift [the poison]	4.45	1.46	1.79	32	.083	0.31

Note. Gender-Stereotypicality: 1 = stereotypically male, 7 = stereotypically female. The noun phrase's translation to English is inserted in parentheses below.

[41] The *t*-test for "das Wasser" [the water] could not be conducted because of its *SD* < 0.01.

Table A11

Verb Phrases Considered Stereotypically Gender-Neutral According to a One-Sample t-Test Against the Scale Midpoint of 4

Verb Phrase	Gender-Stereotypicality					
	M	*SD*	*t*	*df*	*p*	Cohen's *d*
wird eingerichtet [is furnished]	3.64	1.60	-1.31	32	.200	0.23
wird gemietet [is rented]	3.76	0.75	-1.85	32	.073	0.32
wird ausgeschaltet [is switched off]	3.79	0.86	-1.42	32	.165	0.25
wird übersehen [is overlooked]	3.79	0.93	-1.31	32	.198	0.23
wird gefilmt [is filmed]	3.79	0.99	-1.23	32	.228	0.21
wird eingeschaltet [is switched on]	3.82	0.77	-1.48	37	.147	0.24
wird aufgehängt [is hung up]	3.82	1.29	-0.88	37	.385	0.14
wird aufgeschlagen [is flipped open]	3.85	0.87	-1.00	32	.325	0.17
wird bestaunt [is gazed]	3.87	0.96	-0.84	37	.405	0.14
wird gefiltert [is filtered]	3.88	0.55	-1.28	32	.211	0.22
wird besichtigt [is inspected]	3.92	0.94	-0.52	37	.608	0.08
wird gesucht [is searched]	3.92	0.94	-0.52	37	.608	0.08
wird gedreht [is spun]	3.94	0.86	-0.40	32	.690	0.07
wird gedimmt [is dimmed]	3.94	0.93	-0.37	32	.712	0.06
wird geschoben [is pushed]	3.95	1.09	-0.30	37	.767	0.05
wird gewechselt [is changed]	3.97	0.43	-0.37	37	.711	0.06
wird abgestellt [is turned off]	3.97	0.64	-0.26	37	.800	0.04
wird bestellt [is ordered]	3.97	0.89	-0.18	37	.856	0.03
wird abgeschnitten [is cut-off]	3.97	0.95	-0.18	32	.856	0.03
wird gelöscht [is deleted]	3.97	0.95	-0.18	32	.856	0.03
wird zerbrochen [is broken]	3.97	1.05	-0.15	37	.878	0.03
wird entsorgt [is disposed]	3.97	1.26	-0.14	32	.891	0.02
wird bewohnt [is inhabited]	4.00	0.84	< 0.01	37	1.000	< 0.01
wird verschüttet [is spilled]	4.00	1.32	< 0.01	70	1.000	< 0.01
wird gesichtet [is spotted]	4.03	0.37	0.44	37	.661	0.07
wird aufgenommen [is absorbed]	4.03	0.68	0.24	37	.812	0.04

Note. Gender-Stereotypicality: 1 = stereotypically male, 7 = stereotypically female. The verb phrase's translation to English is inserted in parentheses below.

(continued)

Table A11

Verb Phrases Considered Stereotypically Gender-Neutral According to a One-Sample t-Test Against the Scale Midpoint of 4 (continued)

Verb Phrase	Gender-Stereotypicality					
	M	*SD*	*t*	*df*	*p*	Cohen's *d*
wird studiert [is studied]	4.03	0.85	0.19	37	.850	0.03
wird geöffnet [is opened]	4.03	0.92	0.18	37	.860	0.03
wird benutzt [is used]	4.03	1.05	0.15	37	.878	0.03
wird ertastet [is fumbled]	4.06	0.93	0.37	32	.712	0.06
wird verstaut [is stowed]	4.06	1.37	0.26	32	.801	0.04
wird gefüllt [is filled]	4.08	0.85	0.57	37	.571	0.09
wird gesehen [is seen]	4.09	0.46	1.14	32	.263	0.20
wird eröffnet [is inaugurated]	4.09	0.63	0.83	32	.414	0.14
wird hingelegt [is laid down]	4.09	0.72	0.72	32	.475	0.13
wird gerieben [is grated]	4.09	0.80	0.65	32	.521	0.11
wird abgeholt [is picked up]	4.09	0.98	0.53	32	.598	0.09
wird gebraten [is roasted]	4.09	1.47	0.36	32	.724	0.06
wird verschickt [is sent]	4.11	0.89	0.73	37	.473	0.12
wird gepellt [is peeled]	4.13	0.96	0.84	37	.405	0.14
wird weggeschmissen [is thrown away]	4.15	1.06	0.82	32	.419	0.14
wird bestückt [is equipped]	4.15	1.46	0.60	32	.555	0.10
wird beschrieben [is described]	4.16	0.68	1.43	37	.160	0.23
wird vorgezeigt [is shown]	4.18	0.68	1.53	32	.136	0.27
wird gespielt [is played]	4.18	1.10	0.95	32	.351	0.16
wird betrachtet [is watched]	4.21	0.78	1.67	37	.103	0.27
wird besorgt [is gotten]	4.26	1.11	1.47	37	.151	0.24
wird gemieden [is avoided]	4.27	1.21	1.87	70	.066	0.22
wird überreicht [is handed over]	4.30	1.02	1.72	32	.096	0.30
wird zugeschnitten [is cut]	4.30	1.49	1.17	32	.251	0.20
wird angesehen [is watched]	4.39	1.12	2.03	32	.051	0.35

Note. Gender-Stereotypicality: 1 = stereotypically male, 7 = stereotypically female. The verb phrase's translation to English is inserted in parentheses below.

Table A12

Noun Phrases Considered Stereotypically Male According to a One-Sample t-Test Against the Scale Midpoint of 4

Noun Phrase	Gender-Stereotypicality					
	M	*SD*	*t*	*df*	*p*	Cohen's *d*
das Rasiermesser [the razor]	1.39	0.84	-26.25	70	< .001	3.12
das Rasierwasser [the aftershave]	1.42	0.75	-28.98	70	< .001	3.44
das Fußballfeld [the soccer field]	1.72	0.91	-21.05	70	< .001	2.50
das Werkzeug [the tool]	1.73	0.97	-19.69	70	< .001	2.34
das Hemd [the shirt]	1.85	1.04	-17.51	70	< .001	2.08
das Motorrad [the motorcycle]	1.87	0.94	-19.06	70	< .001	2.26
das Bier [the beer]	1.89	0.98	-18.18	70	< .001	2.16
das Stadion [the stadium]	1.94	1.01	-17.11	70	< .001	2.03
das Steak [the steak]	2.03	1.01	-16.39	70	< .001	1.95
das Wettbüro [the betting office]	2.07	0.95	-17.19	70	< .001	2.04
das Quad [the quad bike]	2.17	1.15	-13.46	70	< .001	1.60
das Motorboot [the motorboat]	2.18	1.07	-14.27	70	< .001	1.69
das Fass [the cask]	2.30	1.09	-13.20	70	< .001	1.57
das Auto [the car]	2.42	1.08	-12.33	70	< .001	1.46
das Eisen [the iron]	2.55	1.11	-11.06	70	< .001	1.31
das Bauwerk [the building]	2.68	1.16	-9.65	70	< .001	1.15
das Holz [the wood]	2.70	1.07	-6.96	32	< .001	1.21
das Ventil [the valve]	2.76	1.09	-6.54	32	< .001	1.14
das Funkgerät [the walkie-talkie]	2.79	1.02	-6.80	32	< .001	1.18
das Boot [the boat]	2.84	1.13	-6.33	37	< .001	1.03
das Kabel [the cable]	2.90	1.00	-9.24	70	< .001	1.10
das Raumschiff [the spaceship]	2.95	1.11	-5.83	37	< .001	0.95
das Tor [the gate]	2.95	1.25	-5.19	37	< .001	0.84
das Lenkrad [the steering wheel]	2.96	1.13	-7.79	70	< .001	0.92
das Unternehmen [the company]	3.03	1.10	-7.48	70	< .001	0.89

Note. Gender-Stereotypicality: 1 = stereotypically male, 7 = stereotypically female. The noun phrase's translation to English is inserted in parentheses below.

(continued)

Table A12

Noun Phrases Considered Stereotypically Male According to a One-Sample t-Test Against the Scale Midpoint of 4 (continued)

Noun Phrase	Gender-Stereotypicality					
	M	*SD*	*t*	*df*	*p*	Cohen's *d*
das Messer [the knife]	3.03	1.24	-4.84	37	< .001	0.78
das Schiff [the ship]	3.05	1.09	-5.36	37	< .001	0.87
das Metermaß [the tape measure]	3.07	1.47	-5.34	70	< .001	0.63
das Flugzeug [the plane]	3.16	1.13	-4.60	37	< .001	0.75
das Zelt [the tent]	3.16	1.13	-4.60	37	< .001	0.75
das Nashorn [the rhinoceros]	3.16	1.22	-4.26	37	< .001	0.69
das Dach [the roof]	3.24	0.83	-5.24	32	< .001	0.91
das Mousepad [the mousepad]	3.39	0.83	-4.21	32	< .001	0.73
das Klebeband [the adhesive tape]	3.39	1.00	-3.49	32	.001	0.61
das Thermostat [the thermostat]	3.39	1.09	-3.20	32	.003	0.56
das Saxophon [the saxophone]	3.48	1.15	-2.58	32	.015	0.45
das Gras [the grass]	3.50	0.98	-3.15	37	.003	0.51
das Gebäude [the building]	3.52	0.91	-3.08	32	.004	0.54
das Teleskop [the telescope]	3.55	0.83	-3.33	37	.002	0.54
das UFO [the UFO]	3.55	0.86	-3.20	37	.003	0.52
das Geld [the money]	3.61	0.82	-2.96	37	.005	0.48
das Ticket [the ticket]	3.61	0.86	-2.62	32	.013	0.46
das T-Shirt [the T-shirt]	3.61	1.09	-2.08	32	.046	0.36
das Quadrat [the square]	3.63	0.85	-2.67	37	.011	0.43
das Fahrrad [the bicycle]	3.66	0.88	-2.40	37	.022	0.39

Note. Gender-Stereotypicality: 1 = stereotypically male, 7 = stereotypically female. The noun phrase's translation to English is inserted in parentheses below.

Table A13

Verb Phrases Considered Stereotypically Male According to a One-Sample t-Test Against the Scale Midpoint of 4

Verb Phrase	Gender-Stereotypicality					
	M	*SD*	*t*	*df*	*p*	Cohen's *d*
wird gegrillt [is grilled]	1.90	1.07	-16.51	70	< .001	1.96
wird gezapft [is tapped]	2.04	1.15	-14.32	70	< .001	1.70
wird installiert [is installed]	2.15	0.94	-16.61	70	< .001	1.97
wird geschärft [is sharpened]	2.23	1.10	-13.62	70	< .001	1.62
wird repariert [is repaired]	2.25	1.13	-13.02	70	< .001	1.55
wird angestochen [is broached]	2.30	1.14	-12.61	70	< .001	1.50
wird geschmiedet [is forged]	2.31	1.26	-11.30	70	< .001	1.34
wird eingebaut [is installed]	2.34	0.97	-14.43	70	< .001	1.71
wird gemäht [is mown]	2.46	1.32	-9.81	70	< .001	1.16
wird angeschlossen [is connected]	2.58	1.09	-10.99	70	< .001	1.30
wird gesteuert [is controlled]	2.63	1.00	-8.45	37	< .001	1.37
wird zerstört [is destroyed]	2.65	1.29	-8.84	70	< .001	1.05
wird gefahren [is driven]	2.69	1.17	-9.47	70	< .001	1.12
wird justiert [is adjusted]	2.70	1.02	-10.71	70	< .001	1.27
wird ausgemessen [is measured]	2.76	1.17	-6.08	32	< .001	1.06
wird aufgebaut [is built up]	2.85	1.25	-5.28	32	< .001	0.92
wird erfunden [is invented]	2.95	1.04	-6.25	37	< .001	1.01
wird geschossen [is shot]	3.00	1.34	-4.62	37	< .001	0.75
wird gestrichen [is painted]	3.00	1.40	-6.00	70	< .001	0.71
wird upgedatet [is updated]	3.03	1.13	-5.33	37	< .001	0.86
wird beladen [is loaded]	3.05	1.09	-5.36	37	< .001	0.87
wird reguliert [is regulated]	3.15	1.18	-4.15	32	< .001	0.72
wird aufgestellt [is positioned]	3.21	0.94	-5.21	37	< .001	0.84
wird produziert [is produced]	3.31	0.94	-6.22	70	< .001	0.74
wird errichtet [is raised]	3.32	1.56	-3.66	70	< .001	0.43
wird erhitzt [is heated]	3.37	1.25	-4.29	70	< .001	0.51

Note. Gender-Stereotypicality: 1 = stereotypically male, 7 = stereotypically female. The verb phrase's translation to English is inserted in parentheses below.

(continued)

Table A13

Verb Phrases Considered Stereotypically Male According to a One-Sample t-Test Against the Scale Midpoint of 4 (continued)

Verb Phrase	Gender-Stereotypicality					
	M	*SD*	*t*	*df*	*p*	Cohen's *d*
wird zugedreht [is turned off]	3.39	1.09	-3.20	32	.003	0.56
wird angefertigt [is manufactured]	3.42	1.44	-3.38	70	.001	0.40
wird eingesteckt [is pouched]	3.47	0.95	-3.41	37	.002	0.55
wird verfolgt [is followed]	3.48	1.09	-2.71	32	.011	0.47
wird erkundet [is explored]	3.48	1.20	-2.46	32	.019	0.43
wird verspottet [is mocked]	3.48	1.27	-3.45	70	.001	0.41
wird geschlossen [is closed]	3.50	0.98	-3.15	37	.003	0.51
wird verschoben [is displaced]	3.55	1.09	-2.39	32	.023	0.42
wird entworfen [is designed]	3.55	1.13	-2.44	37	.020	0.40
wird geräumt [is evacuated]	3.59	1.27	-2.71	70	.009	0.32
wird gedruckt [is printed]	3.63	0.93	-3.32	70	.001	0.39
wird gestimmt [is tuned]	3.70	0.77	-2.26	32	.031	0.39
wird betreten [is entered]	3.82	0.51	-2.22	37	.033	0.36

Note. Gender-Stereotypicality: 1 = stereotypically male, 7 = stereotypically female. The verb phrase's translation to English is inserted in parentheses below.

Table A14

Noun Phrases Considered Stereotypically Female According to a One-Sample t-Test Against the Scale Midpoint of 4

Noun Phrase	Gender-Stereotypicality					
	M	*SD*	*t*	*df*	*p*	Cohen's *d*
das Medikament [the medication]	4.18	0.51	2.22	37	.033	0.36
das Müsli [the muesli]	4.24	0.68	2.16	37	.037	0.35
das Bett [the bed]	4.26	0.69	2.37	37	.023	0.38
das Buch [the book]	4.33	0.69	2.77	32	.009	0.48
das Waschbecken [the sink]	4.39	1.10	2.21	37	.034	0.36
das Gefrierfach [the freezer]	4.41	0.89	3.88	70	< .001	0.46
das Kissen [the pillow]	4.42	1.00	2.59	37	.014	0.42
das Herbstfest [the autumn festival]	4.42	1.00	2.44	32	.021	0.42
das Deodorant [the deodorant]	4.45	1.09	2.39	32	.023	0.42
das Poster [the poster]	4.50	0.76	4.04	37	< .001	0.66
das Geschenk [the gift]	4.52	1.03	2.86	32	.007	0.50
das Telefon [the telephone]	4.53	1.13	2.86	37	.007	0.46
das Café [the café]	4.58	0.87	3.81	32	.001	0.66
das Bonbon [the bonbon]	4.61	0.86	4.36	37	< .001	0.71
das Portrait [the portrait]	4.61	0.92	4.07	37	< .001	0.66
das Magazin [the magazine]	4.64	1.19	3.06	32	.004	0.53
das Augenlid [the eyelid]	4.67	0.96	4.00	32	< .001	0.70
das Obst [the fruit]	4.70	0.98	4.07	32	< .001	0.71
das Foto [the photo]	4.74	1.03	4.40	37	< .001	0.71
das Bild [the picture]	4.76	1.00	4.35	32	< .001	0.76
das Familienfest [the family party]	4.88	1.05	4.79	32	< .001	0.83
das Spülbecken [the sink]	4.92	1.00	7.74	70	< .001	0.92
das Ceranfeld [the ceramic hob]	4.99	1.25	6.66	70	< .001	0.79
das Rezept [the recipe]	5.11	1.05	8.93	70	< .001	1.06
das Dessert [the dessert]	5.18	1.06	6.88	37	< .001	1.12
das Geschenkpapier [the gift wrap]	5.18	1.26	5.38	32	< .001	0.94

Note. Gender-Stereotypicality: 1 = stereotypically male, 7 = stereotypically female. The noun phrase's translation to English is inserted in parentheses below.

(continued)

Table A14

Noun Phrases Considered Stereotypically Female According to a One-Sample t-Test Against the Scale Midpoint of 4 (continued)

Noun Phrase	Gender-Stereotypicality					
	M	*SD*	*t*	*df*	*p*	Cohen's *d*
das Spülmittel [the detergent]	5.21	1.11	9.22	70	< .001	1.09
das Porzellan [the porcelain]	5.26	1.03	7.55	37	< .001	1.22
das Sekretariat [the secretariat]	5.27	1.07	10.00	70	< .001	1.19
das Kochfeld [the stove]	5.34	1.10	10.30	70	< .001	1.22
das Kochbuch [the cookbook]	5.45	1.13	10.81	70	< .001	1.28
das Haaröl [the hair oil]	5.49	1.66	7.56	70	< .001	0.90
das Märchenbuch [the fairy-tale book]	5.50	1.08	8.53	37	< .001	1.38
das Pferd [the horse]	5.53	1.29	7.30	37	< .001	1.18
das Baby [the baby]	5.58	1.02	12.99	70	< .001	1.54
das Armband [the bracelet]	5.77	1.06	14.13	70	< .001	1.68
das Schmuckstück [the trinket]	6.14	0.98	18.50	70	< .001	2.19
das Märchenschloss [the fairy castle]	6.39	0.90	22.37	70	< .001	2.65
das Ballett [the ballet]	6.44	0.82	24.93	70	< .001	2.96
das Haarband [the hair band]	6.46	0.86	24.18	70	< .001	2.87
das Puppenhaus [the dollhouse]	6.51	0.73	28.77	70	< .001	3.41
das Make-Up [the make-up]	6.54	0.89	23.96	70	< .001	2.84
das Glätteisen [the hair straightener]	6.62	0.72	30.47	70	< .001	3.62
das Kleid [the dress]	6.66	0.89	25.10	70	< .001	2.98
das Ballkleid [the ball gown]	6.77	0.57	41.34	70	< .001	4.91

Note. Gender-Stereotypicality: 1 = stereotypically male, 7 = stereotypically female. The noun phrase's translation to English is inserted in parentheses below.

Table A15

Verb Phrases Considered Stereotypically Female According to a One-Sample t-Test Against the Scale Midpoint of 4

Verb Phrase	Gender-Stereotypicality					
	M	*SD*	*t*	*df*	*p*	Cohen's *d*
wird zusammengestellt [is compiled]	4.29	0.87	2.06	37	.047	0.33
wird ausgefüllt [is filled out]	4.30	0.53	3.29	32	.002	0.57
wird wahrgenommen [is perceived]	4.32	0.78	2.51	37	.016	0.41
wird ausgewählt [is selected]	4.32	0.81	2.41	37	.021	0.39
wird geholt [is fetched]	4.32	0.96	2.02	37	.050	0.33
wird gelesen [is read]	4.36	0.70	2.99	32	.005	0.52
wird angehört [is listened]	4.39	0.59	4.09	37	< .001	0.66
wird aufgerollt [is coiled]	4.39	0.86	2.62	32	.013	0.46
wird beobachtet [is observed]	4.39	1.03	2.20	32	.035	0.38
wird verlassen [is left]	4.39	1.10	2.21	37	.034	0.36
wird gefunden [is found]	4.42	0.83	3.14	37	.003	0.51
wird aufgehoben [is picked up]	4.45	1.09	2.39	32	.023	0.42
wird abgetaut [is defrosted]	4.51	1.15	3.73	70	< .001	0.44
wird eingepackt [is packed]	4.51	1.15	3.73	70	< .001	0.44
wird abgeheftet [is filed]	4.53	1.11	2.93	37	.006	0.47
wird geführt [is guided]	4.55	1.37	2.49	37	.017	0.40
wird bewundert [is admired]	4.61	0.86	4.03	32	< .001	0.70
wird umgerührt [is stirred up]	4.63	1.13	3.46	37	.001	0.56
wird besucht [is visited]	4.67	0.99	3.87	32	.001	0.67
wird versprüht [is sprayed]	4.70	0.98	4.07	32	< .001	0.71
wird gezeichnet [is drawn]	4.73	0.98	4.28	32	< .001	0.74
wird belegt [is occupied]	4.76	1.02	4.59	37	< .001	0.74
wird aufbewahrt [is stored]	4.79	1.27	3.57	32	.001	0.62
wird vorgelesen [is read aloud]	4.87	1.09	4.89	37	< .001	0.79
wird zubereitet [is prepared]	4.95	1.01	5.77	37	< .001	0.94
wird ausgesucht [is chosen]	5.00	1.14	5.41	37	< .001	0.88

Note. Gender-Stereotypicality: 1 = stereotypically male, 7 = stereotypically female. The verb phrase's translation to English is inserted in parentheses below.

(continued)

Table A15

Verb Phrases Considered Stereotypically Female According to a One-Sample t-Test Against the Scale Midpoint of 4 (continued)

Verb Phrase	Gender-Stereotypicality					
	M	*SD*	*t*	*df*	*p*	Cohen's *d*
wird ausgeschüttelt [is shaken out]	5.05	1.06	6.10	37	< .001	0.99
wird gekauft [is bought]	5.06	1.30	4.70	32	< .001	0.82
wird aufgetragen [is applied]	5.11	1.20	7.80	70	< .001	0.93
wird gemalt [is painted]	5.12	1.02	6.29	32	< .001	1.10
wird gebacken [is baked]	5.13	1.32	5.29	37	< .001	0.86
wird gefärbt [is dyed]	5.15	0.91	7.31	32	< .001	1.27
wird gewiegt [is cradled]	5.20	1.24	8.15	70	< .001	0.97
wird gestreichelt [is stroked]	5.21	1.02	7.33	37	< .001	1.19
wird organisiert [is organized]	5.21	1.32	5.29	32	< .001	0.92
wird desinfiziert [is disinfected]	5.28	1.16	9.30	70	< .001	1.10
wird angelächelt [is smiled at]	5.30	0.98	7.61	32	< .001	1.32
wird aufgeräumt [is tidied]	5.31	1.23	9.00	70	< .001	1.07
wird abgewaschen [is washed up]	5.35	1.04	10.92	70	< .001	1.30
wird gereinigt [is purified]	5.45	1.17	10.46	70	< .001	1.24
wird verschönert [is beautified]	5.45	1.55	7.90	70	< .001	0.94
wird ausgeschmückt [is embellished]	5.58	1.05	12.65	70	< .001	1.50
wird gebügelt [is ironed]	5.70	1.20	11.97	70	< .001	1.42
wird gekocht [is cooked]	5.77	1.03	14.51	70	< .001	1.72
wird geputzt [is cleaned]	5.79	1.11	13.61	70	< .001	1.62
wird dekoriert [is decorated]	5.85	1.01	15.41	70	< .001	1.83
wird genäht [is sewn]	6.28	0.88	21.82	70	< .001	2.59
wird bestickt [is embroidered]	6.45	0.88	23.60	70	< .001	2.80
wird geflochten [is braided]	6.55	0.73	29.32	70	< .001	3.48

Note. Gender-Stereotypicality: 1 = stereotypically male, 7 = stereotypically female. The verb phrase's translation to English is inserted in parentheses below.

A Pretest III – Pretest Instructions

The German wording of the instructions on Pretest III (see Section 4.3) and its translation to English in a paragraph below:

> Liebe Teilnehmerin, lieber Teilnehmer, wir führen eine Vorstudie durch, in der es um die allgemeine Bewertung von Stimmen geht. Hierzu präsentieren wir Ihnen zwei Blöcke mit jeweils 14 gesprochenen Sätzen. Hören Sie sich die einzelnen Sätze eines Blocks bitte nacheinander an. Wenn Sie eine Aufnahme vollständig angehört haben, klicken Sie bitte auf "weiter", um den nächsten Satz abzuspielen. Wenn Sie alle Sätze eines Blocks gehört haben, bitten wir Sie, Fragen zu den Aufnahmen, die Sie soeben gehört haben, zu beantworten. Hören Sie sich anschließend den zweiten Block der Audioaufnahmen an. Es geht einzig um die Bewertung der Stimmen, nicht um den Inhalt der Sätze. Bitte antworten Sie so spontan wie möglich. Es gibt keine richtigen oder falschen Antworten. Zur Bearbeitung steht Ihnen eine siebenstufige Skala zur Verfügung. Markieren Sie bitte das Feld, das am ehesten Ihrer Einschätzung entspricht (1 = Aussage trifft überhaupt nicht zu, 7 = Aussage trifft sehr zu). Die Bearbeitung dieses Fragebogens wird ca. 10 Minuten dauern. Vielen Dank für Ihre Teilnahme!

[Dear participant, we are conducting a pretest on people's general evaluation of speaker voices. Doing so, you will be presented two blocks of recorded sentences. Each block contains 14 single sentences. Please listen carefully to each single sentence. After you have completely listened to a sentence, click "next" to play the next sentence. After you have listened to all sentences of a block, we ask you to answer some questions concerning the recordings you have recently listened to. Then, please listen to the second block of recorded sentences. Answer as spontaneous as possible. There are no right or wrong answers. Please use the 7-point Likert scale to give your responses. Mark that field that represents best your evaluation (1 = statement does not apply at all, 7 = statement fully applies). It might take about 10 minutes to complete the pretest. Thanks for your participation!"]

A Pretest III – Material: List of Sentences Recorded by a Male and a Female Speaker

Table A16

Sentences Used in Pretest III (Section 4.3) to Identify Whether the Speaker Voices Were Perceived as Male and Female

Sentence	NP1 Gender	NP2 Gender
Das Kochfeld wird justiert von dem Mechatroniker. [The stove is adjusted by the mechatronics engineer.]	female	male
Das Steak wird gegrillt von der Hausfrau. [The steak is roasted by the housewife.]	male	female
Das Fußballfeld wird gemäht von dem Türsteher. [The soccer field is mowed by the bouncer.]	male	male
Das Stadion wird geräumt von der Flugbegleiterin. [The stadium was evacuated by the stewardess.]	male	female
Das Ceranfeld wird eingebaut von dem Goldgräber. [The ceramic hob is installed by the gold digger.]	female	male
Das Puppenhaus wird errichtet von der Parfümeriefachverkäuferin. [The doll house is built by the perfumery shop assistant.]	female	female
Das Ballett wird verspottet von dem Metzger. [The ballet is mocked by the butcher.]	female	male
Das Schmuckstück wird geschmiedet von der Floristin. [The jewel is forged by the florist.]	female	female
Das Werkzeug wird zerstört von dem Bauarbeiter. [The tool is destroyed by the builder.]	male	male
Das Motorboot wird gefahren von der Zahnarzthelferin. [The motorboat is driven by the dental assistance.]	male	female
Das Armband wird angefertigt von dem Feuerwehrmann. [The bracelet is manufactured by the firefighter.]	female	male
Das Spülbecken wird installiert von der Kosmetikerin. [The sink is installed by the cosmetician.]	female	female
Das Kabel wird angeschlossen von dem Soldaten. [The cable is connected by the soldier.]	male	male
Das Rasiermesser wird geschärft von der Friseurin. [The razor is sharpened by the hair dresser.]	male	female

Note. NP = Noun Phrase, NP1 Gender/NP2 Gender = gender-stereotypicality of the respective noun phrase. The sentence's translation to English is inserted in parentheses below.

A Experiment 1 (Adverbs – Humans) – Verbal Stimuli: The Experimental Sentences

Table A17

The Experimental Sentences of Experiment 1 in Tuples, Their Classification in Terms of Speaker Voice, Adverb Gender-Stereotypicality, Adverb Connotation, and NP2 Gender, and Their Exact Word On- and Offsets

Tuple	Speaker Voice	Adv. Gender	Adv. Conn.	NP2 Gender	Experimental Sentence	NP1 Off	V1 On	V1 Off	Adv. On	Adv. Off	V2 On	V2 Off	Prep. On	Prep. Off	NP2 On	NP2 Off
1	Male	Male	Pos.	Male	Das Xylophon wird **sachlich** gespielt von dem Bauarbeiter.	1036	1459	1867	2156	2821	3221	3902	4432	4643	4651	5611
	Female				[The xylophone is **objectively** played by the construction worker.]	1243	1459	1755	2156	2882	3282	4088	4806	5184	5342	6626
	Male	Female			Das Xylophon wird **gefühlvoll** gespielt von dem Bauarbeiter.	1128	1459	1910	2126	2971	3333	4002	4470	4615	4615	5579
	Female				[The xylophone is **emotionally** played by the construction worker.]	1132	1459	1805	2156	3239	3639	4336	4698	5171	5396	6788
2	Male	Male	Neg.	Female	Das Ladegerät wird **aggressiv** bestellt von der Zahnarzthelferin.	1022	1560	2061	2315	3094	3494	4149	4712	4977	4979	6245
	Female				[The charger is **aggressively** ordered by the dental assistant.]	1258	1560	1901	2315	3213	3613	4365	5305	5646	5796	7453
	Male	Female			Das Ladegerät wird **nervös** bestellt von der Zahnarzthelferin.	989	1560	2030	2315	2949	3349	4007	4524	4717	4717	5937
	Female				[The charger is **nervously** ordered by the dental assistant.]	1305	1560	1875	2315	3084	3484	4291	4771	5192	5453	7173
3	Male	Male	Neg.	Male	Das Blatt wird **prahlerisch** zugeschnitten von dem Feuerwehrmann.	639	1182	1638	2045	2620	2996	3903	4334	4465	4466	5418
	Female				[The sheet of paper is **boastfully** cut by the firefighter.]	798	1210	1527	2065	2981	3381	4217	4595	4989	5226	6397
	Male	Female			Das Blatt wird **panisch** zugeschnitten von dem Feuerwehrmann.	637	1210	1640	2065	2596	2996	3905	4375	4499	4499	5418
	Female				[The sheet of paper is **panically** cut by the firefighter.]	759	1210	1495	2065	2706	3106	4016	4493	4969	5151	6333
4	Male	Male	Pos.	Female	Das Radiergummi wird **fachmännisch** genutzt von der Kosmetikerin.	968	1393	1898	2161	2998	3441	4055	4610	4756	4756	5843
	Female				[The eraser is **professionally** used by the cosmetician.]	1151	1393	1701	2161	3063	3872	4589	4859	5295	5338	6754
	Male	Female			Das Radiergummi wird **fingerfertig** genutzt von der Kosmetikerin.	1009	1393	1882	2161	3074	3688	4322	4849	4965	4965	6114
	Female				[The eraser is **dexterously** used by the cosmetician.]	1115	1393	1720	2161	3121	3836	4586	4799	5270	5485	7164

Note. On/Off = On- and offsets: the time in ms from sentence onset. The onset of NP1 was 0ms for all sentences and is thus not listed here for clarity of presentation. Adv./NP2 Gender = Adverb/NP2 Gender-Stereotypicality, Adv. Conn. = Adverb Connotation, Pos. = Positive, Neg. Negative, NP = Noun Phrase, V = Verb, Adv. = Adverb (marked in bold), Prep. = Preposition. The sentence's translation to English is inserted in parentheses below.

(continued)

Table A17

The Experimental Sentences of Experiment 1 in Tuples, Their Classification in Terms of Speaker Voice, Adverb Gender-Stereotypicality, Adverb Connotation, and NP2 Gender, and Their Exact Word On- and Offsets (continued)

Tuple	Speaker Voice	Adv. Gender	Adv. Conn.	NP2 Gender	Experimental Sentence	NP1 Off	V1 On	V1 Off	Adv. On	Adv. Off	V2 On	V2 Off	Prep. On	Prep. Off	NP2 On	NP2 Off
5	Male	Male	Pos.	Male	Das Paket wird **gelassen** verschickt von dem Soldaten.	844	1316	1841	2176	2809	3209	3889	4561	4686	4721	5623
	Female				[The parcel is **calmly** sent by the soldier.]	1009	1392	1701	2176	2913	3313	4164	4710	5142	5468	6752
	Male	Female			Das Paket wird **vorsorglich** verschickt von dem Soldaten.	863	1392	1872	2176	3073	3421	4069	4572	4746	4746	5718
	Female				[The parcel is **preventively** sent by the soldier.]	896	1392	1702	2176	3191	3541	4237	4763	5181	5261	6599
6	Male	Male	Neg.	Female	Das Plakat wird **cholerisch** aufgehängt von der Friseurin.	906	1411	1912	2209	2875	3275	4062	4388	4570	4575	5415
	Female				[The poster is **irascibly** hung-up by the hair dresser.]	1086	1411	1693	2209	3070	3470	4363	4627	5012	5164	6127
	Male	Female			Das Plakat wird **ungeschickt** aufgehängt von der Friseurin.	929	1411	1955	2209	2927	3327	4065	4688	4784	4792	5626
	Female				[The poster is **clumsily** hung-up by the hair dresser.]	1099	1411	1761	2209	3089	3489	4394	5066	5533	5609	6848
7	Male	Male	Neg.	Male	Das Ei wird **gierig** gepellt von dem Mechatroniker.	765	1220	1709	1913	2492	2881	3482	3972	4149	4152	5206
	Female				[The egg is **greedily** peeled by the mechatronics engineer.]	747	1220	1510	1913	2614	3014	3814	4295	4732	4924	6301
	Male	Female			Das Ei wird **naiv** gepellt von dem Mechatroniker.	707	1220	1696	1913	2530	2881	3476	3873	4009	4009	5036
	Female				[The egg is **naively** peeled by the mechatronics engineer.]	790	1220	1550	1913	2840	3119	3926	4605	4980	5123	6459
8	Male	Male	Pos.	Female	Das Portemonnaie wird **pragmatisch** gesucht von der Hausfrau.	1115	1608	2069	2401	3243	3643	4323	4730	4864	4918	5692
	Female				[The wallet is **pragmatically** searched by the housewife.]	1161	1608	1924	2401	3364	3764	4451	4699	5160	5265	6227
	Male	Female			Das Portemonnaie wird **zuverlässig** gesucht von der Hausfrau.	995	1608	2069	2401	3371	3771	4437	4925	5117	5137	5924
	Female				[The wallet is **reliably** searched by the housewife.]	1198	1608	1921	2401	3334	3734	4501	5019	5477	5773	6859
9	Male	Male	Pos.	Male	Das Tier wird **gekonnt** gemieden von dem Türsteher.	752	1250	1743	2158	2807	3207	3844	4440	4630	4631	5486
	Female				[The animal is **skillfully** avoided by the bouncer.]	812	1250	1596	2158	2929	3329	4069	4345	4802	4903	5962
	Male	Female			Das Tier wird **rücksichtsvoll** gemieden von dem Türsteher.	774	1250	1765	2158	3042	3442	4097	4535	4756	4757	5567
	Female				[The animal is **considerately** avoided by the bouncer.]	879	1250	1674	2158	3241	3641	4402	4983	5413	5562	6687

Note. On/Off = On- and offsets: the time in ms from sentence onset. The onset of NP1 was 0ms for all sentences and is thus not listed here for clarity of presentation. Adv./NP2 Gender = Adverb/NP2 Gender-Stereotypicality, Adv. Conn. = Adverb Connotation, Pos. = Positive, Neg. Negative, NP = Noun Phrase, V = Verb, Adv. = Adverb (marked in bold), Prep. = Preposition. The sentence's translation to English is inserted in parentheses below.

(continued)

Table A17

The Experimental Sentences of Experiment 1 in Tuples, Their Classification in Terms of Speaker Voice, Adverb Gender-Stereotypicality, Adverb Connotation, and NP2 Gender, and Their Exact Word On- and Offsets (continued)

Tuple	Speaker Voice	Adv. Gender	Adv. Conn.	NP2 Gender	Experimental Sentence	NP1 Off	V1 On	V1 Off	Adv. On	Adv. Off	V2 On	V2 Off	Prep. On	Prep. Off	NP2 On	NP2 Off
10	Male	Male	Neg.	Female	Das Handy wird **gefühllos** ausgeschaltet von der Flugbegleiterin.	780	1201	1673	2000	2847	3247	4195	4539	4703	4704	5876
	Female				[The cell phone is **callously** turned-off by the stewardess.]	809	1201	1460	2000	3001	3401	4420	4723	5163	5235	6602
	Male	Female			Das Handy wird **heimtückisch** ausgeschaltet von der Flugbegleiterin.	739	1201	1694	2000	2797	3197	4130	4425	4584	4585	4639
	Female				[The cell phone is **insidiously** turned-off by the stewardess.]	860	1201	1577	2000	2799	3199	4273	4681	5127	5166	6511
11	Male	Male	Neg.	Male	Das Dokument wird **angeberisch** gelöscht von dem Goldgräber.	979	1533	2011	2350	3188	3588	4193	4687	4812	4813	5756
	Female				[The document is **boastfully** deleted by the gold digger.]	1113	1533	1819	2350	3434	3834	4554	5075	5502	5673	6794
	Male	Female			Das Dokument wird **penibel** gelöscht von dem Goldgräber.	912	1533	2002	2350	3000	3400	4008	4520	4682	4683	5571
	Female				[The document is **meticulously** deleted by the gold digger.]	1065	1533	1852	2350	3173	3573	4222	4904	5377	5396	6521
12	Male	Male	Pos.	Female	Das Brötchen wird **spontan** besorgt von der Parfümeriefachverkäuferin.	825	1308	1711	1993	2756	3156	3825	4384	4542	4543	6283
	Female				[The roll is **spontaneously** gotten by the perfumery shop assistant.]	1017	1308	1579	1993	2869	3269	4006	4194	4669	4727	7035
	Male	Female			Das Brötchen wird **euphorisch** besorgt von der Parfümeriefachverkäuferin.	852	1308	1752	1993	2748	3148	3777	4452	4650	4651	6285
	Female				[The roll is **euphorically** gotten by the perfumery shop assistant.]	952	1308	1627	1993	2894	3294	3945	4190	4582	4721	6937
13	Male	Male	Pos.	Male	Das Radio wird **mutig** abgestellt von dem Metzger.	896	1332	1835	2101	2731	3131	3930	4511	4660	4661	5379
	Female				[The radio is **bravely** turned off by the butcher.]	1033	1332	1680	2101	2818	3218	4140	4566	4975	5018	5961
	Male	Female			Das Radio wird **feinfühlig** abgestellt von dem Metzger.	945	1332	1808	2101	2943	3343	4250	4650	4861	4862	5560
	Female				[The radio is **sensitively** turned off by the butcher.]	1014	1332	1686	2101	3061	3461	4310	4765	5244	5329	6321
14	Male	Male	Neg.	Female	Das Aquarium wird **leichtsinnig** gefüllt von der Floristin.	1075	1464	1966	2243	3064	3464	4091	4516	4736	4737	5582
	Female				[The aquarium is **recklessly** filled by the florist.]	1174	1464	1807	2243	3106	3506	4225	4621	5051	5052	6167
	Male	Female			Das Aquarium wird **intrigant** gefüllt von der Floristin.	1019	1464	1901	2243	3037	3437	3997	4565	4781	4782	5609
	Female				[The aquarium is **schemingly** filled by the florist.]	1094	1464	1768	2243	3160	3560	4205	4695	5078	5079	6307

Note. On/Off = On- and offsets: the time in ms from sentence onset. The onset of NP1 was 0ms for all sentences and is thus not listed here for clarity of presentation. Adv./NP2 Gender = Adverb/NP2 Gender-Stereotypicality, Adv. Conn. = Adverb Connotation, Pos. = Positive, Neg. Negative, NP = Noun Phrase, V = Verb, Adv. = Adverb (marked in bold), Prep. = Preposition. The sentence's translation to English is inserted in parentheses below.

(continued)

Table A17

The Experimental Sentences of Experiment 1 in Tuples, Their Classification in Terms of Speaker Voice, Adverb Gender-Stereotypicality, Adverb Connotation, and NP2 Gender, and Their Exact Word On- and Offsets (continued)

Tuple	Speaker Voice	Adv. Gender	Adv. Conn.	NP2 Gender	Experimental Sentence	NP1 Off	V1 On	V1 Off	Adv. On	Adv. Off	V2 On	V2 Off	Prep. On	Prep. Off	NP2 On	NP2 Off
15	Male	Male	Neg.	Male	Das Publikum wird **bedenkenlos** betrachtet von dem Bauarbeiter.	1077	1459	1888	2142	3081	3481	4261	4823	5014	5015	5934
	Female				[The audience is **regardlessly** viewed by the construction worker.]	1075	1459	1791	2142	3209	3609	4461	5104	5539	5586	6896
	Male	Female			Das Publikum wird **verlogen** betrachtet von dem Bauarbeiter.	1034	1459	1904	2142	2925	3325	4076	4632	4813	4814	5769
	Female				[The audience is **mendaciously** watched by the construction worker.]	1088	1459	1796	2142	2930	3330	4214	4770	5189	5223	6592
16	Male	Male	Pos.	Female	Das Zimmer wird **strategisch** eingerichtet von der Zahnarzthelferin.	798	1271	1660	1981	2946	3346	4181	4831	5065	5066	6245
	Female				[The room is **strategically** furnished by the dental assistant.]	796	1271	1617	1981	2963	3363	4312	4585	5090	5110	6761
	Male	Female			Das Zimmer wird **stilvoll** eingerichtet von der Zahnarzthelferin.	950	1271	1792	1981	2914	3314	4269	4771	4958	4959	6314
	Female				[The room is **stylishly** furnished by the dental assistant.]	807	1271	1614	1981	2872	3272	4289	4896	5382	5393	6992
17	Male	Male	Pos.	Male	Das Museum wird **stolz** eröffnet von dem Feuerwehrmann.	914	1259	1801	2025	2730	3130	3875	4494	4704	4705	5696
	Female				[The museum is **proudly** inaugurated by the firefighter.]	945	1259	1664	1968	2704	3104	3864	4368	4825	4886	6101
	Male	Female			Das Museum wird **wortgewandt** eröffnet von dem Feuerwehrmann.	914	1259	1760	2028	2877	3277	4011	4535	4751	4752	5791
	Female				[The museum is **eloquently** inaugurated by the firefighter.]	982	1259	1706	1941	3087	3487	4274	4641	5166	5177	6510
18	Male	Male	Neg.	Female	Das Handtuch wird **rücksichtslos** gewechselt von der Kosmetikerin.	886	1252	1825	1999	3001	3401	4230	4729	4960	4961	6125
	Female				[The towel is **ruthlessly** changed by the cosmetician.]	859	1315	1648	1999	3151	3551	4380	4547	4885	4886	6206
	Male	Female			Das Handtuch wird **verträumt** gewechselt von der Kosmetikerin.	821	1252	1782	1999	2867	3267	4074	4370	4558	4559	5653
	Female				[The towel is **dreamily** changed by the cosmetician.]	843	1310	1659	1999	2942	3427	4203	4643	5170	5171	6667
19	Male	Male	Neg.	Male	Das Gift wird **unvorsichtig** entsorgt von dem Soldaten.	844	1199	1768	2116	3138	3538	4290	4578	4876	4877	5800
	Female				[The poison is **carelessly** disposed by the soldier.]	914	1269	1668	2165	3406	3806	4741	4987	5482	5483	6882
	Male	Female			Das Gift wird **übervorsichtig** entsorgt von dem Soldaten.	785	1252	1768	2198	3247	3745	4499	4955	5243	5244	6207
	Female				[The poison is **overcautiously** disposed by the soldier.]	834	1199	1621	2116	3486	3886	4936	4985	5541	5542	6937

Note. On/Off = On- and offsets: the time in ms from sentence onset. The onset of NP1 was 0ms for all sentences and is thus not listed here for clarity of presentation. Adv./NP2 Gender = Adverb/NP2 Gender-Stereotypicality, Adv. Conn. = Adverb Connotation, Pos. = Positive, Neg. Negative, NP = Noun Phrase, V = Verb, Adv. = Adverb (marked in bold), Prep. = Preposition. The sentence's translation to English is inserted in parentheses below.

(continued)

Table A17

The Experimental Sentences of Experiment 1 in Tuples, Their Classification in Terms of Speaker Voice, Adverb Gender-Stereotypicality, Adverb Connotation, and NP2 Gender, and Their Exact Word On- and (continued)

Tuple	Speaker Voice	Adv. Gender	Adv. Conn.	NP2 Gender	Experimental Sentence	NP1 Off	V1 On	V1 Off	Adv. On	Adv. Off	V2 On	V2 Off	Prep. On	Prep. Off	NP2 On	NP2 Off
20	Male	Male	Pos.	Female	Das Album wird **unerschrocken** aufgeschlagen von der Friseurin.	966	1259	1863	2068	2974	3541	4462	4726	4988	4989	5976
	Female				[The album is **fearlessly** flipped open by the hair dresser.]	1011	1259	1653	2162	3099	3570	4682	4829	5226	5227	6396
	Male	Female			Das Album wird **gewissenhaft** aufgeschlagen von der Friseurin.	965	1312	1931	2156	3158	3558	4493	4867	5078	5079	6112
	Female				[The album is **conscientiously** flipped open by the hair dresser.]	977	1316	1625	2144	3261	3661	4778	5317	5840	5841	7089
21	Male	Male	Pos.	Male	Das Fischfilet wird **überzeugt** gebraten von dem Mechatroniker.	1134	1461	1925	2218	2976	3376	4193	4315	4631	4632	5704
	Female				[The fish fillet is **confidently** roasted by the mechatronics engineer.]	1274	1461	1794	2218	3117	3665	4544	5011	5464	5465	6866
	Male	Female			Das Fischfilet wird **hingebungsvoll** gebraten von dem Mechatroniker.	968	1461	1908	2218	3320	3720	4492	4806	4986	4987	6105
	Female				[The fish fillet is **dedicatedly** roasted by the mechatronics engineer.]	1173	1461	1773	2218	3412	3812	4714	5061	5595	5678	7165
22	Male	Male	Neg.	Female	Das Spielzeug wird **unordentlich** überreicht von der Hausfrau.	963	1414	1919	2223	3165	3565	4324	4667	4822	4823	5701
	Female				[The toy is **untidily** handed over by the housewife.]	1078	1414	1818	2223	3357	3757	4630	4825	5277	5354	6479
	Male	Female			Das Spielzeug wird **hinterlistig** überreicht von der Hausfrau.	916	1414	1903	2223	3132	3532	4314	4708	4939	4940	5748
	Female				[The toy is **cunningly** handed over by the housewife.]	1212	1414	1886	2223	3313	3713	4701	5027	5499	5600	6725
23	Male	Male	Neg.	Male	Das Labyrinth wird **respektlos** besichtigt von dem Türsteher.	1031	1489	2031	2317	3338	3738	4591	5085	5295	5296	6277
	Female				[The labyrinth is **disrespectfully** viewed by the bouncer.]	1119	1489	1866	2317	3437	3837	4706	5204	5744	5745	6890
	Male	Female			Das Labyrinth wird **orientierungslos** besichtigt von dem Türsteher.	1015	1489	2000	2317	3495	3895	4725	5238	5441	5442	6372
	Female				[The labyrinth is **disorientedly** viewed by the bouncer.]	1134	1489	1958	2317	3882	4282	5177	5537	6035	6036	7342
24	Male	Male	Pos.	Female	Das Naturereignis wird **heldenhaft** bestaunt von der Flugbegleiterin.	1276	1838	2345	2647	3415	3815	4494	4796	4938	4939	6227
	Female				[The natural event is **heroically** gazed by the stewardess.]	1541	1838	2172	2647	3520	3920	4700	4961	5500	5628	7224
	Male	Female			Das Naturereignis wird **aufmerksam** bestaunt von der Flugbegleiterin.	1303	1838	2350	2647	3635	4035	4723	5294	5532	5533	6694
	Female				[The natural event is **attentively** gazed by the stewardess.]	1733	1838	2213	2647	3630	4030	4874	5130	5619	5724	7260

Note. On/Off = On- and offsets: the time in ms from sentence onset. The onset of NP1 was 0ms for all sentences and is thus not listed here for clarity of presentation. Adv./NP2 Gender = Adverb/NP2 Gender-Stereotypicality, Adv. Conn. = Adverb Connotation, Pos. = Positive, Neg. Negative, NP = Noun Phrase, V = Verb, Adv. = Adverb (marked in bold), Prep. = Preposition. The sentence's translation to English is inserted in parentheses below.

(continued)

Table A17

The Experimental Sentences of Experiment 1 in Tuples, Their Classification in Terms of Speaker Voice, Adverb Gender-Stereotypicality, Adverb Connotation, and NP2 Gender, and Their Exact Word On- and Offsets (continued)

Tuple	Speaker Voice	Adv. Gender	Adv. Conn.	NP2 Gender	Experimental Sentence	NP1 Off	V1 On	V1 Off	Adv. On	Adv. Off	V2 On	V2 Off	Prep. On	Prep. Off	NP2 On	NP2 Off
25	Male	Male	Pos.	Male	Das Haus wird **professionell** gemietet von dem Goldgräber.	790	1460	1941	2574	3629	4029	4726	5354	5557	5559	6582
	Female				[The house is **professionally** rented by the gold digger.]	978	1460	1830	2574	3732	4132	4928	5111	5582	5585	6776
	Male	Female			Das Haus wird **ordnungsgemäß** gemietet von dem Goldgräber.	793	1460	1957	2574	3642	4042	4738	5346	5538	5541	6458
	Female				[The house is **properly** rented by the gold digger.]	933	1460	1768	2574	3865	4265	5054	5300	5788	5838	7108
26	Male	Male	Neg.	Female	Das Klavier wird **verantwortungslos** vorgezeigt von der Parfümeriefach-verkäuferin.	921	1478	1911	2344	3592	3992	4901	5202	5365	5367	7176
	Female				[The piano is **irresponsibly** shown by the perfumery shop assistant.]	1157	1478	1932	2368	4098	4498	5504	5707	6135	6202	8637
	Male	Female			Das Klavier wird **überempfindlich** vorgezeigt von der Parfümeriefachver-käuferin.	916	1478	2049	2368	3419	3819	4787	5162	5372	5373	7131
	Female				[The piano is **hypersensitively** shown by the perfumery shop assistant.]	1021	1478	1860	2368	3680	4080	5046	5305	5677	5743	8204
27	Male	Male	Neg.	Male	Das Opossum wird **unzuverlässig** gefilmt von dem Metzger.	956	1447	2021	2327	3379	3779	4421	4795	5038	5039	5760
	Female				[The opossum is **unreliably** filmed by the butcher.]	1068	1447	1811	2327	3533	3933	4769	4993	5495	5531	6427
	Male	Female			Das Opossum wird **eingebildet** gefilmt von dem Metzger.	929	1447	2033	2327	3237	3637	4270	4825	5046	5047	5749
	Female				[The opossum is **conceitedly** filmed by the butcher.]	1132	1447	1795	2327	3495	3895	4614	4759	5174	5175	6022
28	Male	Male	Pos.	Female	Das Wasser wird **stark** gefiltert von der Floristin.	785	1416	1969	2158	2705	3105	3840	4523	4705	4706	5541
	Female				[The water is **strongly** filtered by the florist.]	907	1416	1704	2158	2803	3203	4086	4284	4757	4773	5875
	Male	Female			Das Wasser wird **sensibel** gefiltert von der Floristin.	791	1416	1873	2158	2945	3345	4090	4635	4829	4830	5711
	Female				[The water is **sensitively** filtered by the florist.]	885	1416	1684	2158	3091	3491	4339	4610	5058	5125	6256
29	Male	Male	Pos.	Male	Das Sofa wird **selbstbewusst** geschoben von dem Bauarbeiter.	844	1398	1863	2236	3200	3600	4248	4737	4944	4945	5834
	Female				[The sofa is **confidently** pushed by the construction worker.]	909	1398	1750	2236	3392	3792	4646	4960	5421	5517	6850
	Male	Female			Das Sofa wird **wohlüberlegt** geschoben von dem Bauarbeiter.	770	1398	1940	2236	3220	3620	4300	4942	5150	5151	6084
	Female				[The sofa is **deliberately** pushed by the construction worker.]	1011	1398	1786	2236	3405	3805	4643	5040	5513	5651	6977

Note. On/Off = On- and offsets: the time in ms from sentence onset. The onset of NP1 was 0ms for all sentences and is thus not listed here for clarity of presentation. Adv./NP2 Gender = Adverb/NP2 Gender-Stereotypicality, Adv. Conn. = Adverb Connotation, Pos. = Positive, Neg. Negative, NP = Noun Phrase, V = Verb, Adv. = Adverb (marked in bold), Prep. = Preposition. The sentence's translation to English is inserted in parentheses below.

(continued)

Table A17

The Experimental Sentences of Experiment 1 in Tuples, Their Classification in Terms of Speaker Voice, Adverb Gender-Stereotypicality, Adverb Connotation, and NP2 Gender, and Their Exact Word On- and Offsets (continued)

Tuple	Speaker Voice	Adv. Gender	Adv. Conn.	NP2 Gender	Experimental Sentence	NP1 Off	V1 On	V1 Off	Adv. On	Adv. Off	V2 On	V2 Off	Prep. On	Prep. Off	NP2 On	NP2 Off
30	Male	Male	Neg.	Female	Das Eichhörnchen wird **gedankenlos** gesichtet von der Zahnarzthelferin.	1074	1666	2170	2512	3556	3956	4718	5231	5466	5467	6670
	Female				[The squirrel is **thoughtlessly** spotted by the dental assistant.]	1289	1666	2065	2512	3740	4140	5010	5525	6009	6010	7665
	Male	Female			Das Eichhörnchen wird **zurückhaltend** gesichtet von der Zahnarzthelferin.	1154	1666	2170	2512	3566	3966	4728	5190	5395	5396	6603
	Female				[The squirrel is **unobtrusively** spotted by the dental assistant.]	1204	1666	2060	2512	3652	4052	4795	5146	5576	5691	7421
31	Male	Male	Neg.	Male	Das Glücksrad wird **grobmotorisch** gedreht von dem Feuerwehrmann.	915	1412	1800	2307	3270	3670	4330	5365	5576	5577	6575
	Female				[The prize-wheel is **grossly** spun by the firefighter.]	1120	1412	1810	2307	3509	3909	4585	4755	5154	5203	6477
	Male	Female			Das Glücksrad wird **selbstzweifelnd** gedreht von dem Feuerwehrmann.	922	1412	1893	2307	3490	3890	4485	5153	5390	5391	6386
	Female				[The prize-wheel is **self-doubtingly** spun by the firefighter.]	1140	1412	1840	2307	3624	4024	4675	4815	5329	5397	6613
32	Male	Male	Pos.	Female	Das Brettspiel wird **risikobereit** angesehen von der Kosmetikerin.	1096	1492	1945	2256	3309	3709	4487	4865	5056	5057	6163
	Female				[The board game is **risk-takingly** watched by the cosmetician.]	1324	1492	1880	2256	3569	3969	4967	5098	5600	5685	7131
	Male	Female			Das Brettspiel wird **fasziniert** angesehen von der Kosmetikerin.	974	1492	1945	2256	3188	3588	4384	4902	5138	5139	6209
	Female				[The board game is **fascinatedly** watched by the cosmetician.]	1154	1492	1865	2256	3465	3865	4710	4794	5211	5283	6733

Note. On/Off = On- and offsets: the time in ms from sentence onset. The onset of NP1 was 0ms for all sentences and is thus not listed here for clarity of presentation. Adv./NP2 Gender = Adverb/NP2 Gender-Stereotypicality, Adv. Conn. = Adverb Connotation, Pos. = Positive, Neg. Negative, NP = Noun Phrase, V = Verb, Adv. = Adverb (marked in bold), Prep. = Preposition. The sentence's translation to English is inserted in parentheses below.

A Experiment 1 (Adverbs – Humans) – Verbal Stimuli: The Filler Sentences

Table A18

The Filler Sentences of Experiment 1, Their Classifications in Terms of NP1, Adverb, Main Verb, and NP2 Gender-Stereotypicality, the NP2's Grammatical Gender, and the Corresponding Questions Related to Either the Content of the Filler Sentences or to the Pictures

Item	NP1 Gender	Adv. Conn.	Main Verb Gender	NP2 Gender	Filler Sentence	Question following the filler sentence	Correct Answer	Question's Focus
F1	Male	Neg.	Male	Male	Das Auto wird unbeholfen repariert von dem Mechatroniker. [The car is clumsily repaired by the mechatronics engineer (male form).]	Trug der Mechatroniker blaue Kleidung? [Was the mechatronics engineer (male form) dressed in blue?]	Yes	Picture
F2	Male	Pos.	Male	Female	Das Steak wird kompetent gegrillt von der Hausfrau. [The steak is competently roasted by the housewife (female form).]	Wird das Steak kompetent gegrillt von der Hausfrau? [Was the steak competently roasted by the housewife (female form)?]	Yes	Sentence
F3	Male	Pos.	Female	Male	Das Zelt wird zukunftsorientiert aufbewahrt von der Mechatronike**rin**. [The tent is future-orientedly repaired by the mechatronics engineer **(female form)**.]	War das Zelt gelb? [Was the tent yellow?]	No	Picture
F4	Male	Neg.	Female	Female	Das Lenkrad wird oberflächlich desinfiziert von dem Haus**mann**. [The steering wheel is superficially disinfected by the house**husband (male form)**.]	Wird das Lenkrad oberflächlich desinfiziert von dem Haus**mann**? [Is the steering wheel superficially disinfected by the house**husband (male form)**?]	Yes	Sentence
F5	Female	Pos.	Male	Male	Das Haaröl wird durchdacht produziert von dem Türsteher. [The hair oil is sophisticatedly produced by the bouncer (male form).]	Trug der Türsteher einen Ledermantel? [Did the bouncer (male form) wear a leather coat?]	No	Picture
F6	Female	Neg.	Female	Female	Das Pferd wird unbeholfen gestreichelt von der Flugbegleiterin. [The horse is clumsily stroked by the stewardess (female form).]	Wird das Pferd unbeholfen gefüttert von der Flugbegleiterin? [Is the horse clumsily fed by the stewardess (female form)?]	No	Sentence

Note. Adv./NP1/Main Verb/NP2 Gender = NP1 Object/Main Verb/NP2 Gender-Stereotypicality, Adv. Conn. = Adverb Connotation, Pos. = Positive, Neg. Negative. Deviations between a NP2 profession's gender-stereotypicality and its grammatical gender are marked in bold. Adverbs used in the filler sentences were stereotypically gender-neutral. The sentence's and the question's translation to English are inserted in parentheses below.

(continued)

Table A18

The Filler Sentences of Experiment 1, Their Classifications in Terms of NP1, Adverb, Main Verb, and NP2 Gender-Stereotypicality, the NP2's Grammatical Gender, and the Corresponding Questions Related to Either the Content of the Filler Sentences or to the Pictures (continued)

Item	NP1 Gender	Adv. Conn.	Main Verb Gender	NP2 Gender	Filler Sentence	Question following the filler sentence	Correct Answer	Question's Focus
F7	Female	Pos.	Female	Male	Das Haarband wird konzentriert geflochten von der Türsteher**in**. [The hair band is focusedly braided by the bouncer **(female form)**.]	War das Haarband grün? [Was the hair band green?]	No	Picture
F8	Female	Neg.	Female	Female	Das Spülmittel wird oberflächlich abgewaschen von dem Flugbeglei**ter** [The detergent is superficially washed up by the steward **(male form)**.]	Wird das Spülmittel oberflächlich abgewaschen von dem Flugbegleit**er**? [Is the detergent superficially washed up by the steward **(male form)**?]	Yes	Sentence
F9	Male	Neg.	Male	Male	Das Fass wird tollpatschig angestochen von dem Goldgräber. [The cask is clumsily broached by the gold digger (male form).]	War das Fass aus Holz? [Was the cask wooden?]	Yes	Picture
F10	Male	Neg.	Male	Female	Das Bier wird ungeduldig gezapft von der Parfümeriefachverkäuferin. [The beer is impatiently tapped by the perfumery shop assistant (female form).]	Wird das Bier ungeduldig gezapft von dem Fris**eur**? [Is the beer impatiently tapped by the hair dresser **(male form)**?]	No	Sentence
F11	Male	Pos.	Female	Male	Das Motorrad wird akribisch verschönert von der Goldgräber**in**. [The motorcycle is meticulously beautified by the gold digger **(female form.)**]	Wird das Motorrad schnell gefahren von der Goldgräber**in**? [Is the motorcycle fastly driven by the gold digger **(female form)**?]	No	Sentence
F12	Male	Pos.	Male	Female	Das Eisen wird routiniert erhitzt von dem Parfümeriefachverkäuf**er**. [The iron is experienced heated by the perfumery shop assistant **(male form)**.]	Trug die Parfümeriefachverkäuferin ein oranges Top? [Did the perfumery shop assistant (female form) wear an orange top?]	Yes	Picture
F13	Female	Pos.	Male	Male	Das Ballett wird kritisch verspottet von dem Metzger. [The ballet is critically mocked by the butcher (male form).]	Wird das Ballett lauthals verspottet von dem Handwerker? [Is the ballet loudly mocked by the craftsman (male form)?]	No	Sentence
F14	Female	Pos.	Male	Female	Das Schmuckstück wird konzentriert geschmiedet von der Floristin. [The jewel is focusedly forged by the florist (female form).]	War das abgebildete Schmuckstück eine Kette? [Was the depicted jewel a necklace?]	No	Picture

Note. Adv./NP1/Main Verb/NP2 Gender = NP1 Object/Main Verb/NP2 Gender-Stereotypicality, Adv. Conn. = Adverb Connotation, Pos. = Positive, Neg. Negative. Deviations between a NP2 profession's gender-stereotypicality and its grammatical gender are marked in bold. Adverbs used in the filler sentences were stereotypically gender-neutral. The sentence's and the question's translation to English are inserted in parentheses below.

(continued)

Table A18

The Filler Sentences of Experiment 1, Their Classifications in Terms of NP1, Adverb, Main Verb, and NP2 Gender-Stereotypicality, the NP2's Grammatical Gender, and the Corresponding Questions Related to Either the Content of the Filler Sentences or to the Pictures (continued)

Item	NP1 Gender	Adv. Conn.	Main Verb Gender	NP2 Gender	Filler Sentence	Question following the filler sentence	Correct Answer	Question's Focus
F15	Female	Pos.	Female	Male	Das Kleid wird geschickt genäht von der Metzger**in**. [The dress is skillfully sewn by the butcher **(female form)**.]	War das Kleid blau? [Was the dress blue?]	Yes	Picture
F16	Female	Neg.	Female	Female	Das Porzellan wird genervt geputzt von dem Floris**ten**. [The porcelain is annoyedly cleaned by the florist **(male form)**.]	Wird das Porzellan genervt geputzt von dem Florist**en**? [Is the porcelain annoyedly cleaned by the florist **(male form)**?]	Yes	Sentence
F17	Male	Neg.	Male	Male	Das Werkzeug wird tollpatschig zerstört von dem Bauarbeiter. [The tool is clumsily destroyed by the construction worker (male form).]	War das abgebildete Werkzeug ein Hammer? [Was the depicted tool a hammer?]	No	Picture
F18	Male	Neg.	Male	Female	Das Motorboot wird langsam gefahren von der Zahnarzthelferin. [The motorboat is slowly driven by the dental assistant (female form).]	Wird das Motorboot langsam gefahren von dem Zahnarzt? [Is the motorboat slowly driven by the dentist (male form)?]	No	Sentence
F19	Male	Neg.	Female	Male	Das Metermaß wird ungeduldig eingepackt von der Bauarbeiter**in**. [The tape measure is impatiently packed by the construction worker **(female form)**.]	Trug die Bauarbeiter**in** einen Helm? [Did the construction worker (female form) wear a helmet?]	Yes	Picture
F20	Male	Neg.	Female	Female	Das Raumschiff wird inkompetent gereinigt von dem Zahnarzthelf**er**. [The spaceship is incompetently cleaned by the dental assistant **(male form)**.]	Wird das Raumschiff inkompetent gereinigt von der Zahnarzt-helferin? [Is the spaceship incompetently cleaned by the dental assistant (female form)?]	No	Sentence
F21	Female	Pos.	Male	Male	Das Armband wird geschickt angefertigt von dem Feuerwehrmann. [The bracelet is skillfully manufactured by the firefighter (male form).]	Wird das Holzspiel geschickt angefertigt von dem Feuerwehr-mann? [Is the wooden toy skillfully manufactured by the firefighter (male form)?]	No	Sentence
F22	Female	Neg.	Male	Female	Das Spülbecken wird spöttisch installiert von der Kosmetikerin. [The sink is mockingly installed by the cosmetician (female form).]	Hatte die Kosmetikerin eine Schleife im Haar? [Did the cosmetician (female form) wear a bow in her hair?]	Yes	Picture

Note. Adv./NP1/Main Verb/NP2 Gender = NP1 Object/Main Verb/NP2 Gender-Stereotypicality, Adv. Conn. = Adverb Connotation, Pos. = Positive, Neg. Negative. Deviations between a NP2 profession's gender-stereotypicality and its grammatical gender are marked in bold. Adverbs used in the filler sentences were stereotypically gender-neutral. The sentence's and the question's translation to English are inserted in parentheses below.

(continued)

Table A18

The Filler Sentences of Experiment 1, Their Classifications in Terms of NP1, Adverb, Main Verb, and NP2 Gender-Stereotypicality, the NP2's Grammatical Gender, and the Corresponding Questions Related to Either the Content of the Filler Sentences or to the Pictures (continued)

Item	NP1 Gender	Adv. Conn.	Main Verb Gender	NP2 Gender	Filler Sentence	Question following the filler sentence	Correct Answer	Question's Focus
F23	Female	Neg.	Female	Male	Das Baby wird ungläubig gewiegt von der Feuerwehr**frau**. [The baby is incredulously cradled by the firefighter **(female form)**.]	Wird das Baby geduldig gefüttert von der Feuerwehr**frau**? [Is the baby patiently fed by the firefighter **(female form)**?]	No	Sentence
F24	Female	Pos.	Female	Female	Das Ballkleid wird inspirierend bestickt von dem Kosmetik**er**. [The ball dress is inspiringly embroidered by the cosmetician **(male form).**]	War das Ballkleid weiß? [Was the ball dress white?]	Yes	Picture
F25	Male	Neg.	Male	Male	Das Kabel wird inkompetent angeschlossen von dem Soldaten. [The cable is incompetently connected by the soldier (male form).]	Wird das Kabel ungeduldig angeschlossen von dem Bäcker? [Is the cable impatiently connected by the baker (male form)?]	No	Sentence
F26	Male	Pos.	Male	Female	Das Rasiermesser wird routiniert geschärft von der Friseurin. [The razor is experiencedly sharpened by the hair dresser (female form).]	War das Rasiermesser aufgeklappt? [Was the razor flipped open?]	Yes	Picture
F27	Male	Pos.	Female	Male	Das Unternehmen wird akribisch dekoriert von der Soldat**in**. [The company is meticulously decorated by the soldier **(female form).**]	Wird das Unternehmen akribisch dekoriert von der Soldat**in**? [Is the company meticulously decorated by the soldier **(female form)**?]	Yes	Sentence
F28	Male	Neg.	Female	Female	Das Hemd wird genervt gebügelt von dem Frise**ur**. [The shirt is annoyedly ironed by the hair dresser **(male form).**]	War das Hemd rot? [Was the shirt red?]	No	Picture
F29	Female	Pos.	Male	Male	Das Kochfeld wird kritisch justiert von dem Mechatroniker. [The stove is critically adjusted by the mechatronics engineer (male form).]	Wird das Kochfeld kritisch ausgebaut von dem Mechatroniker? [Is the stove critically removed by the mechatronics engineer (male form)?]	No	Sentence
F30	Female	Neg.	Male	Female	Das Kochbuch wird ungläubig gedruckt von der Hausfrau. [The cookbook is incredulously printed by the housewife (female form).]	Hielt die Hausfrau einen Besen in der Hand? [Did the housewife (female form) hold a broom in her hand?]	Yes	Picture
F31	Female	Neg.	Female	Male	Das Gefrierfach wird langsam abgetaut von der Mechatroniker**in**. [The freezer is slowly defrosted by the mechatronics engineer **(female form).**]	Wird das Gefrierfach langsam abgetaut von dem Mechatroniker? [Is the freezer slowly defrosted by the mechatronics engineer (male form)?]	No	Sentence

Note. Adv./NP1/Main Verb/NP2 Gender = NP1 Object/Main Verb/NP2 Gender-Stereotypicality, Adv. Conn. = Adverb Connotation, Pos. = Positive, Neg. Negative. Deviations between a NP2 profession's gender-stereotypicality and its grammatical gender are marked in bold. Adverbs used in the filler sentences were stereotypically gender-neutral. The sentence's and the question's translation to English are inserted in parentheses below.

(continued)

Table A18

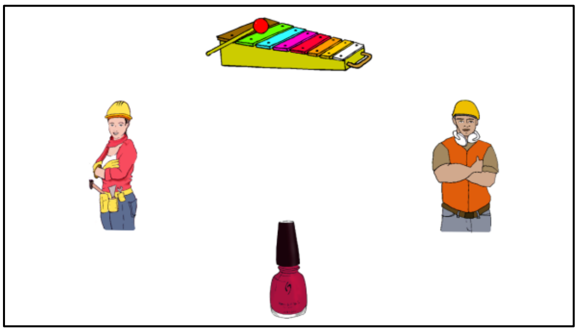
The Filler Sentences of Experiment 1, Their Classifications in Terms of NP1, Adverb, Main Verb, and NP2 Gender-Stereotypicality, the NP2's Grammatical Gender, and the Corresponding Questions Related to Either the Content of the Filler Sentences or to the Pictures (continued)

Item	NP1 Gender	Adv. Conn.	Main Verb Gender	NP2 Gender	Filler Sentence	Question following the filler sentence	Correct Answer	Question's Focus
F32	Female	Pos.	Female	Female	Das Rezept wird durchdacht gekocht von dem Haus**mann**. [The receipt is sophisticatedly cooked by the househusband **(male form).**]	Wird das Rezept durchdacht gekocht von der Hausfrau? [Is the receipt sophisticatedly cooked by the housewife (female form)?]	No	Sentence
F33	Male	Pos.	Male	Male	Das Fußballfeld wird zukunftsorierntiert gemäht von dem Türsteher. [The soccer field is future-orientedly mown by the bouncer (male form).]	War das Fußballfeld blau? [Was the soccer field blue?]	No	Picture
F34	Male	Pos.	Male	Female	Das Stadium wird pünktlich geräumt von der Flugbegleiterin. [The stadium is punctually evacuated by the stewardess (female form).]	War im Stadion eine amerikanische Flagge abgebildet? [Was there an American flag depicted in the stadium?]	Yes	Picture
F35	Male	Pos.	Female	Male	Das Bauwerk wird inspirierend ausgeschmückt von der Türsteher**in**. [The building is inspiringly embellished by the bouncer **(female form)**.]	Wird das Haus inspirierend ausgeschmückt von der Türsteher**in**? [Was the house inspiringly embellished by the bouncer **(female form)**?]	No	Sentence
F36	Male	Neg.	Female	Female	Das Teleskop wird spöttisch bewundert von dem Flugbegleit**er**. [The telescope is mockingly admired by the steward **(male form)**.]	Hatte das Teleskop drei Standbeine? [Did the telescope stand on three pillars?]	Yes	Picture
F37	Female	Pos.	Male	Male	Das Bett wird kompetent eingebaut von dem Goldgräber. [The bed is competently installed by the gold digger (male form).]	War die Bettwäsche rot und weiß kariert? [Was the linen red and white checkered?]	Yes	Picture
F38	Female	Pos.	Male	Female	Das Puppenhaus wird flink errichtet von der Parfümeriefachverkäu-ferin. [The doll house is swiftly raised by the perfumery shop assistant (fe-male form).]	Wird das Puppenhaus langsam errichtet von der Parfümeriefach-verkäuferin? [Is the doll house slowly raised by the perfumery shop assistant (female form)?]	No	Sentence
F39	Female	Pos.	Female	Male	Das Sekretariat wird pünktlich aufgeräumt von der Goldgräber**in**. [The secretariat is punctually tidied by the gold digger **(female form)**.]	War auf dem Bild des Sekretariats ein Computerbildschirm zu sehen? [Was there a pc screen in the depicted secretariat?]	Yes	Picture
F40	Female	Pos.	Female	Female	Das Make-Up wird flink aufgetragen von dem Parfümeriefachver-käuf**er**. [The make-up is swiftly applied by the perfumery shop assistant **(male form)**.]	Wird das Make-Up ungeschickt aufgetragen von dem Parfü-meriefachverkäuf**er**? [Is the make-up clumsily applied by the perfumery shop assistant **(male form)**?]	No	Sentence

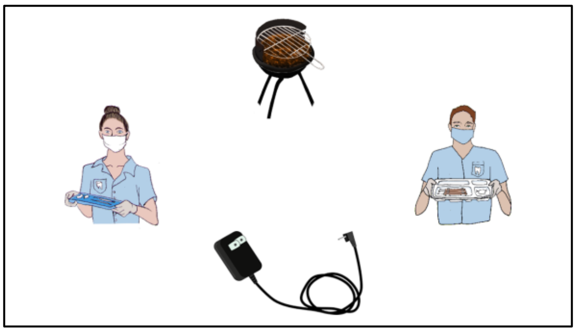
Note. Adv./NP1/Main Verb/NP2 Gender = NP1 Object/Main Verb/NP2 Gender-Stereotypicality, Adv. Conn. = Adverb Connotation, Pos. = Positive, Neg. Negative. Deviations between a NP2 profession's gender-stereotypicality and its grammatical gender are marked in bold. Adverbs used in the filler sentences were stereotypically gender-neutral. The sentence's and the question's translation to English are inserted in parentheses below.

A Experiment 1 (Adverbs – Humans) – Visual Stimuli: The Experimental Stimuli

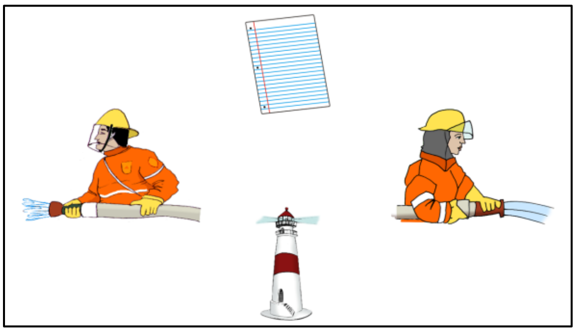
Each of the visual scenes below represents the content of the two sentences within an experimental tuple. The number of the tuple the visual scene refers to is indicated below (see Appendix A, Table A17 for the full set of experimental sentences).



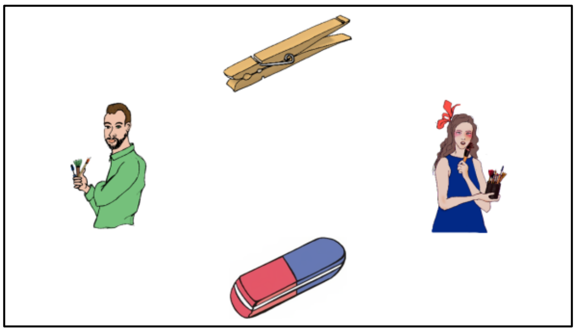
Tuple 1



Tuple 2

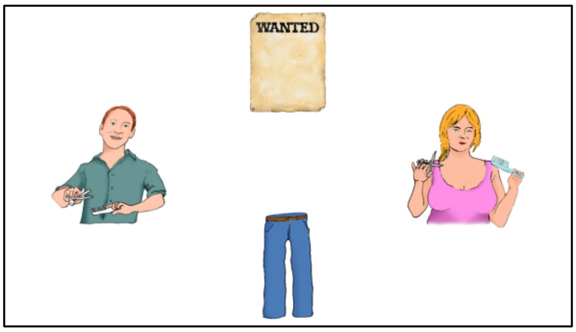


Tuple 3



Tuple 4

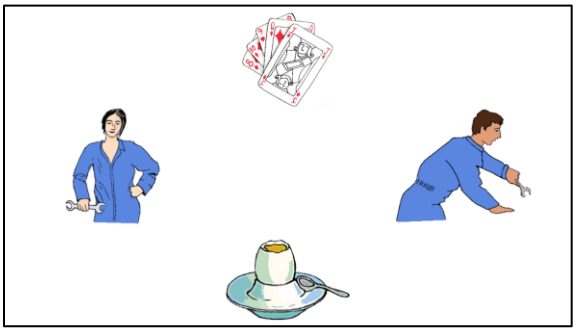
Tuple 5



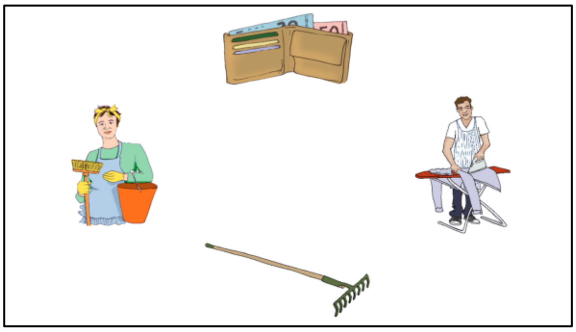
Tuple 6

(continued)

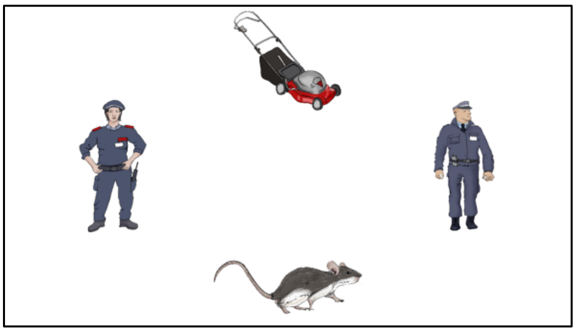
Each of the visual scenes below represent the content of the two sentences within an experimental tuple. The number of the tuple the visual scene refers to is indicated below (see Appendix A, Table A17 for the full set of experimental sentences) (continued).



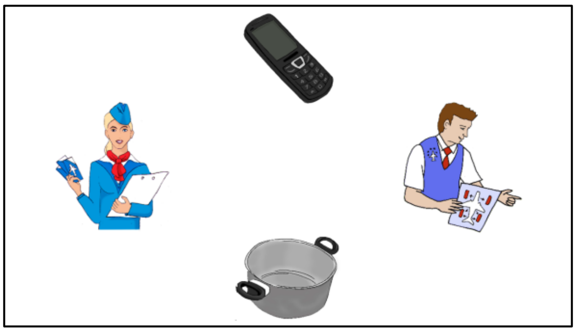
Tuple 7



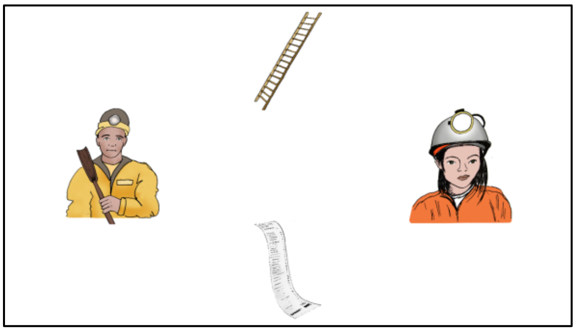
Tuple 8



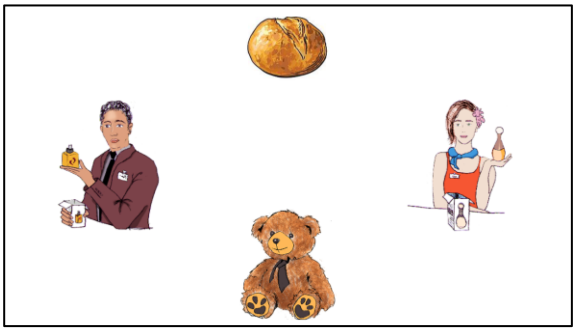
Tuple 9



Tuple 10



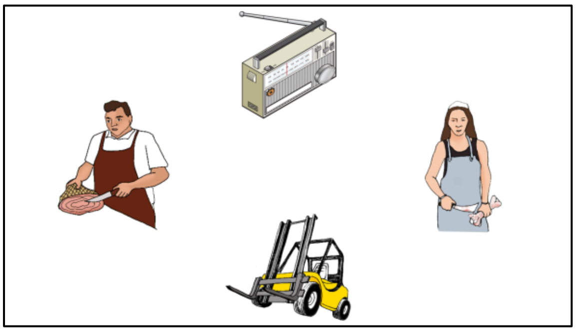
Tuple 11



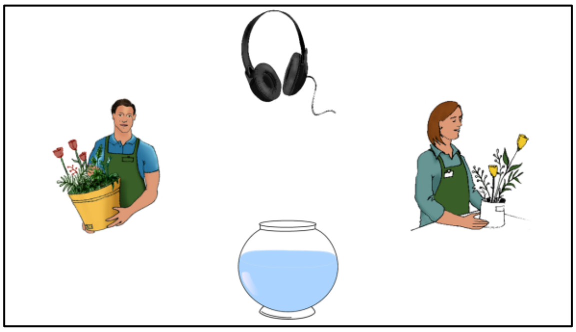
Tuple 12

(continued)

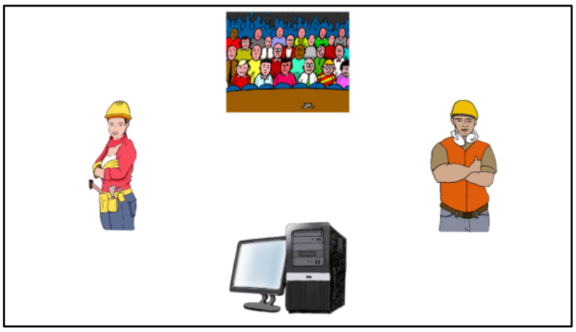
Each of the visual scenes below represent the content of the two sentences within an experimental tuple. The number of the tuple the visual scene refers to is indicated below (see Appendix A, Table A17 for the full set of experimental sentences) (continued).



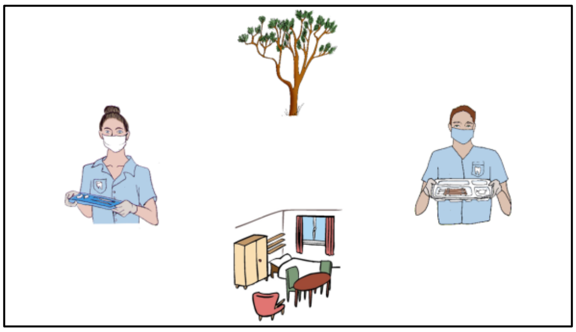
Tuple 13



Tuple 14

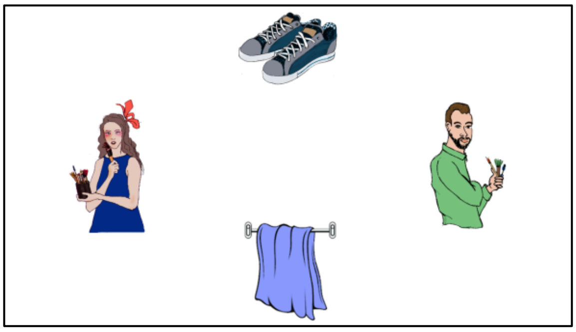


Tuple 15



Tuple 16

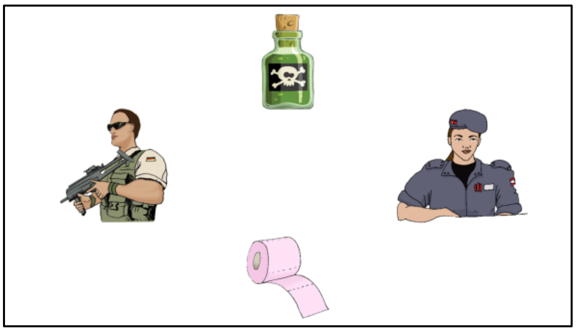
Tuple 17



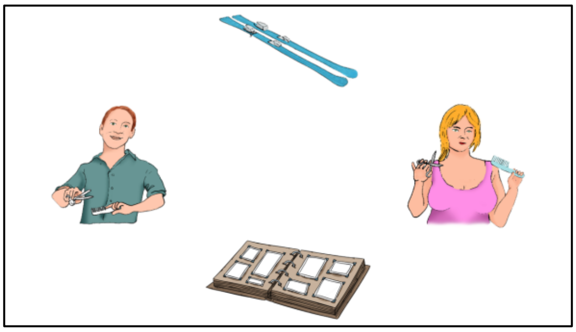
Tuple 18

(continued)

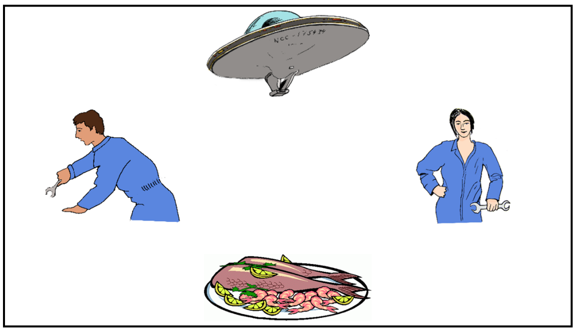
Each of the visual scenes below represent the content of the two sentences within an experimental tuple. The number of the tuple the visual scene refers to is indicated below (see Appendix A, Table A17 for the full set of experimental sentences) (continued).



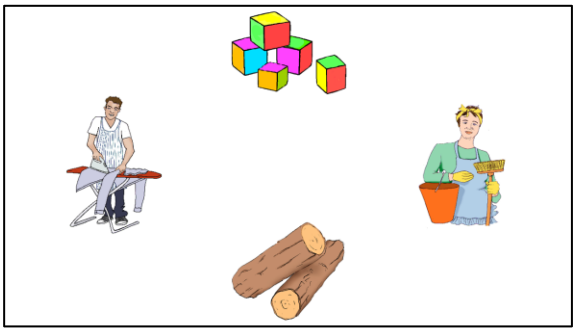
Tuple 19



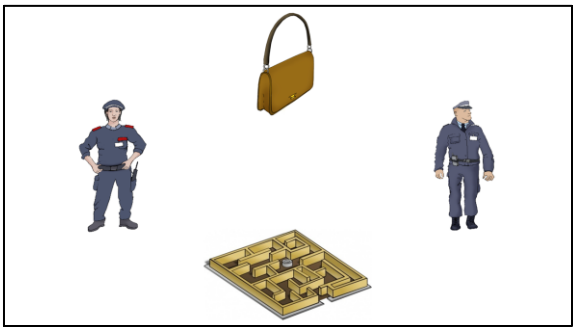
Tuple 20



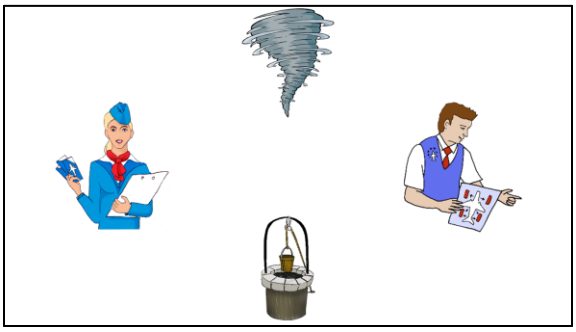
Tuple 21



Tuple 22



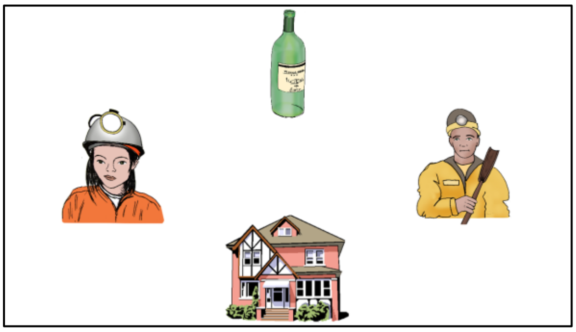
Tuple 23



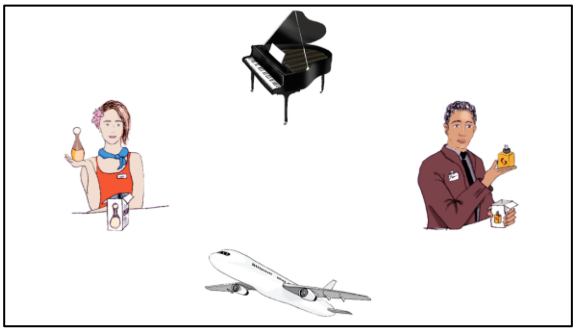
Tuple 24

(continued)

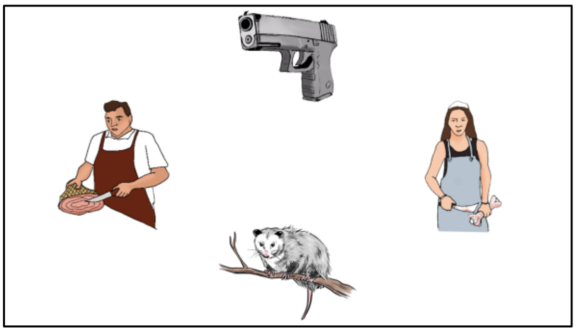
Each of the visual scenes below represent the content of the two sentences within an experimental tuple. The number of the tuple the visual scene refers to is indicated below (see Appendix A, Table A17 for the full set of experimental sentences) (continued).



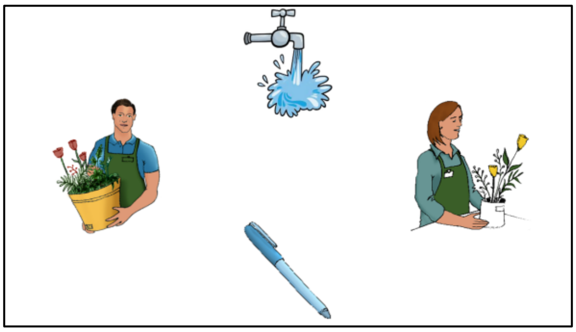
Tuple 25



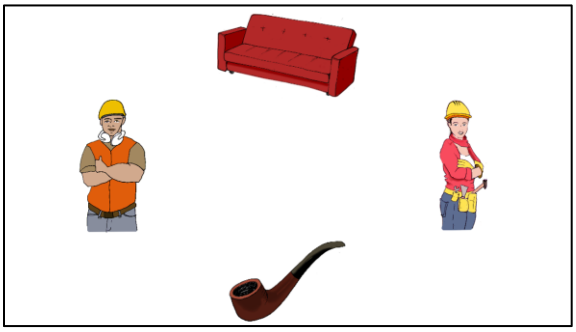
Tuple 26



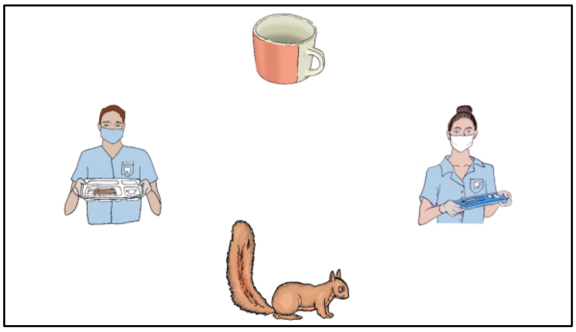
Tuple 27



Tuple 28



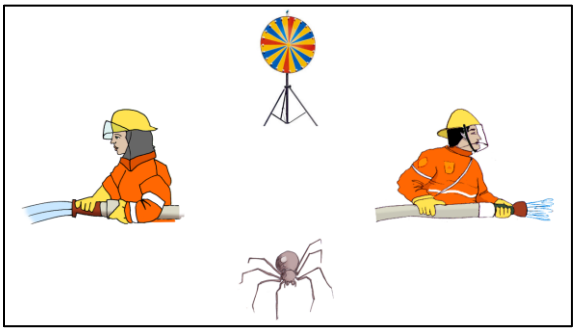
Tuple 29



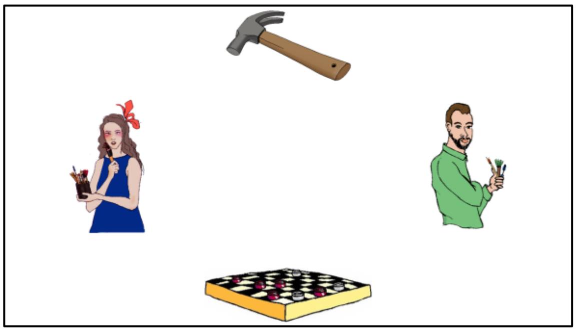
Tuple 30

(continued)

Each of the visual scenes below represent the content of the two sentences within an experimental tuple. The number of the tuple the visual scene refers to is indicated below (see Appendix A, Table A17 for the full set of experimental sentences) (continued).



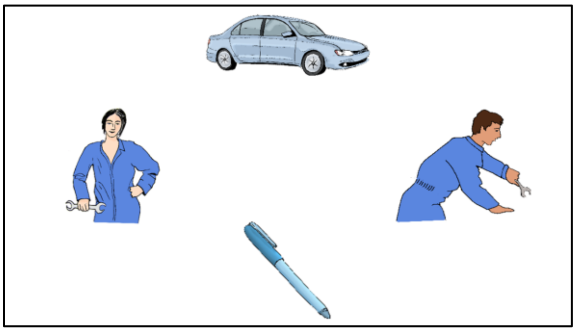
Tuple 31



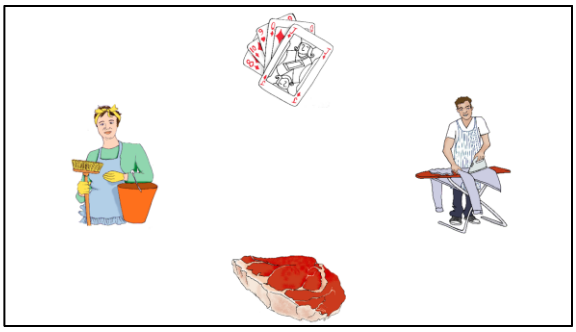
Tuple 32

A Experiment 1 (Adverbs – Humans) – Visual Stimuli: The Fillers

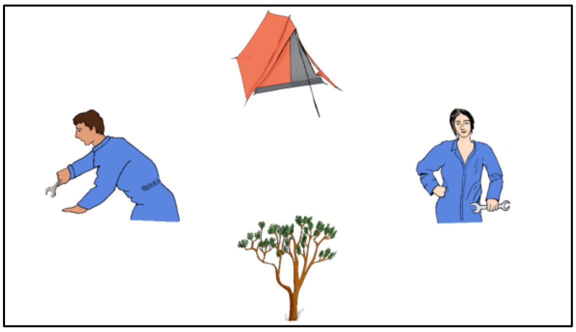
Each of the visual scenes below represents the content of a filler sentence. The number of the filler sentence a visual scene refers to is indicated below (see Appendix A, Table A18 for the full set of filler sentences).



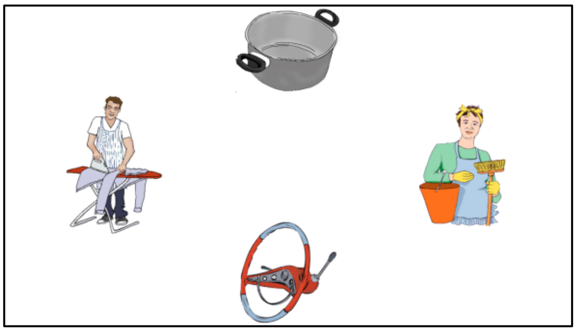
F1



F2

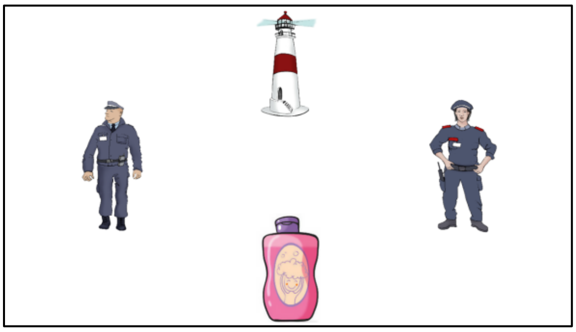


F3

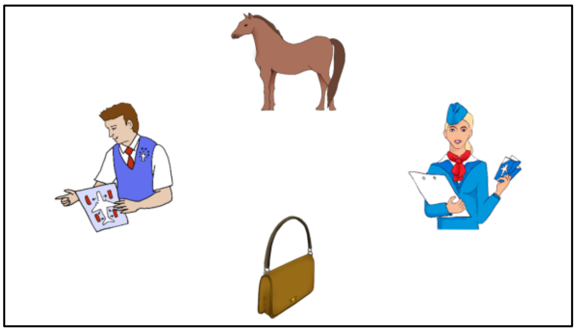


F4

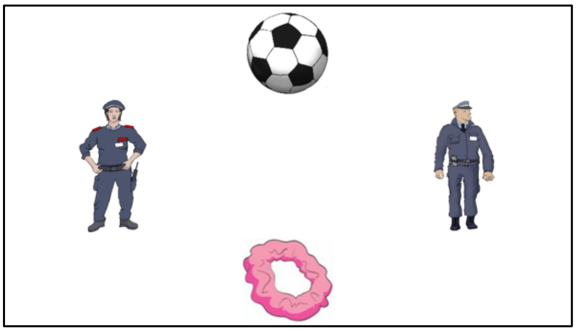
Each of the visual scenes below represents the content of a filler sentence. The number of the filler sentence a visual scene refers to is indicated below (see Appendix A, Table A18 for the full set of filler sentences) (continued).



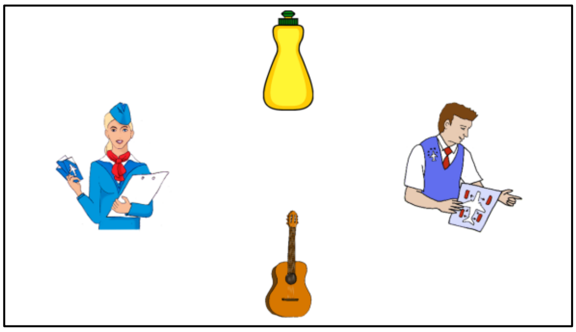
F5



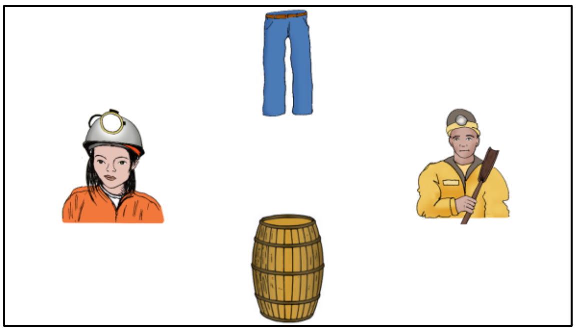
F6



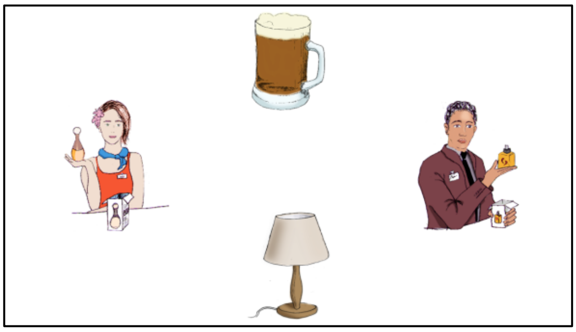
F7



F8

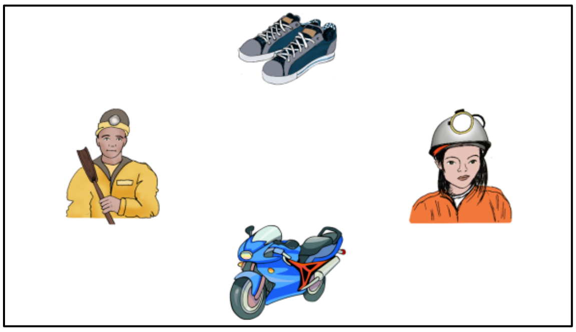


F9

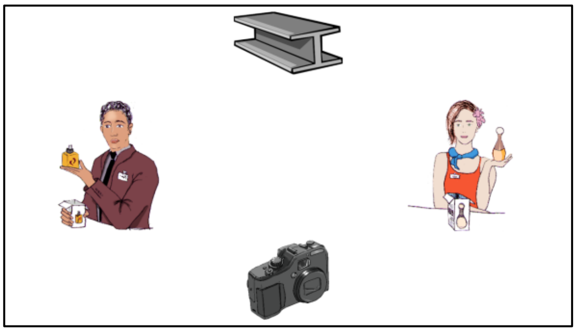


F10

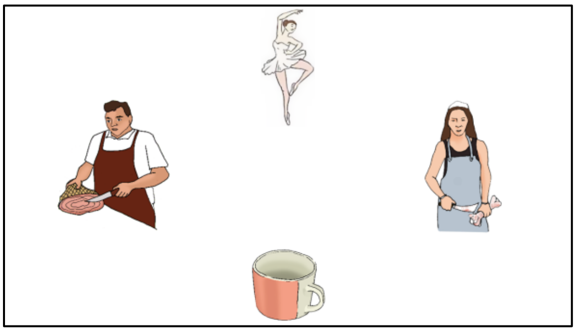
Each of the visual scenes below represents the content of a filler sentence. The number of the filler sentence a visual scene refers to is indicated below (see Appendix A, Table A18 for the full set of filler sentences) (continued).



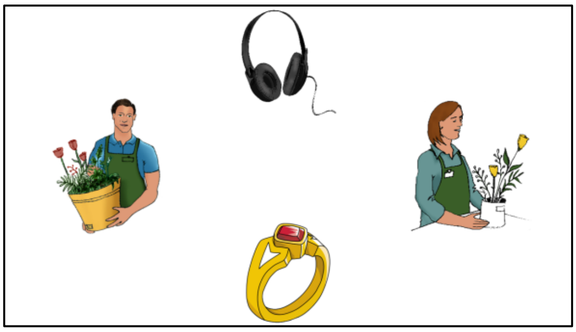
F11



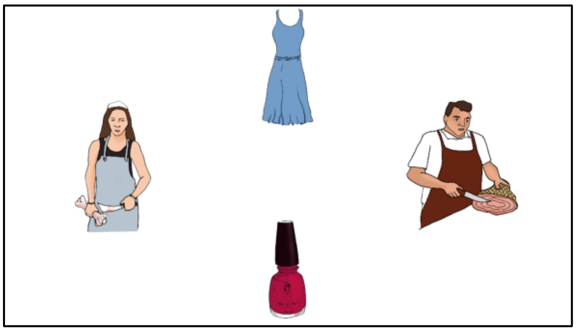
F12



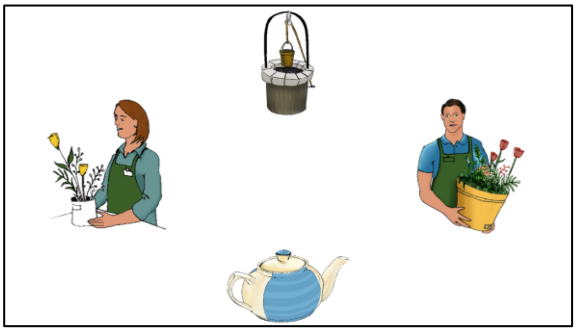
F13



F14

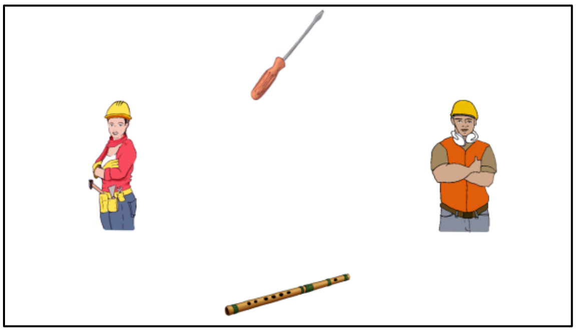


F15

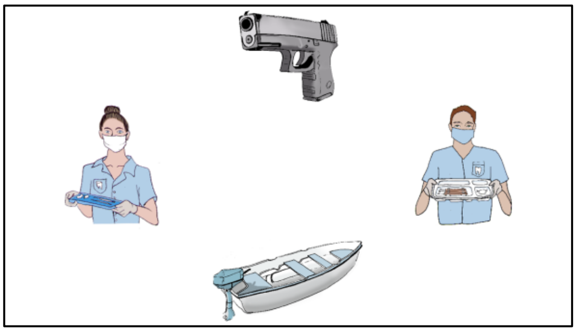


F16

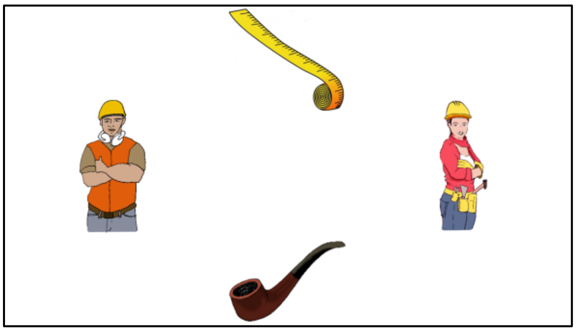
Each of the visual scenes below represents the content of a filler sentence. The number of the filler sentence a visual scene refers to is indicated below (see Appendix A, Table A18 for the full set of filler sentences) (continued).



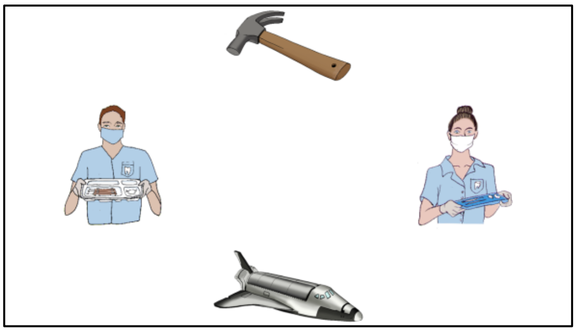
F17



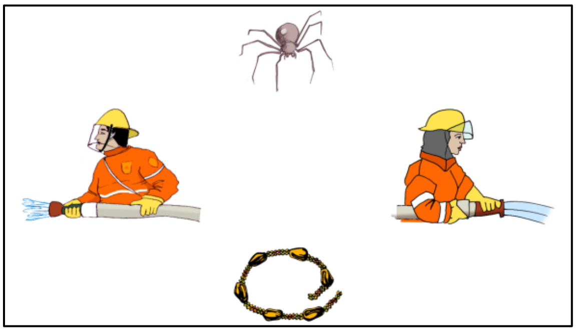
F18



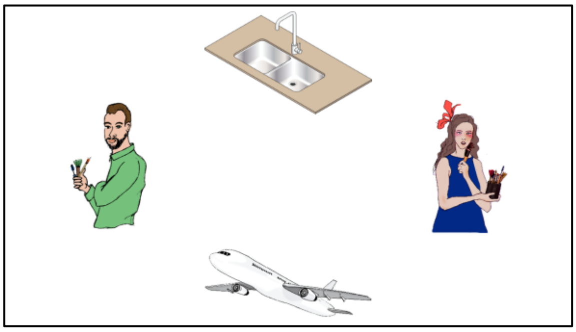
F19



F20

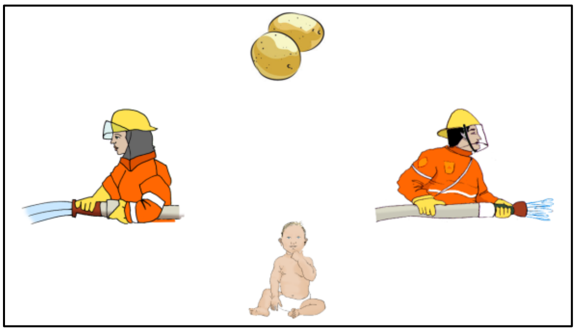


F21

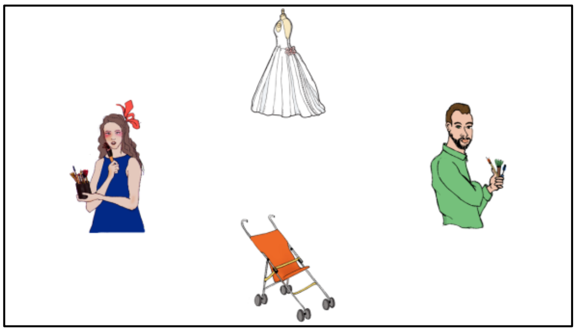


F22

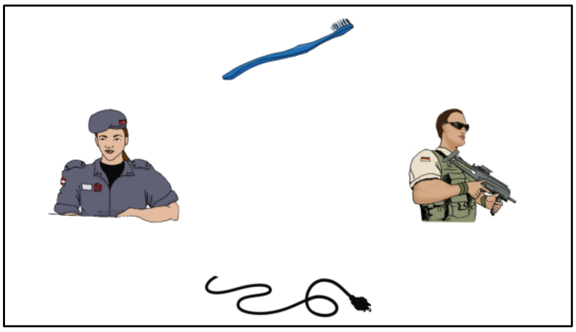
Each of the visual scenes below represents the content of a filler sentence. The number of the filler sentence a visual scene refers to is indicated below (see Appendix A, Table A18 for the full set of filler sentences) (continued).



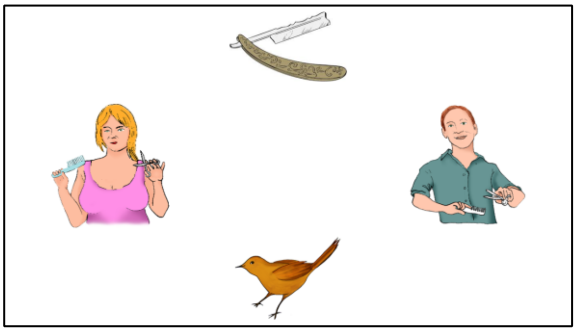
F23



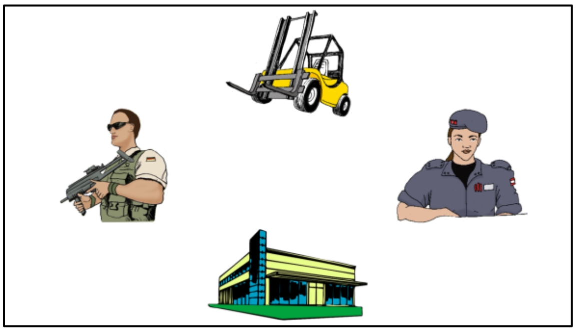
F24



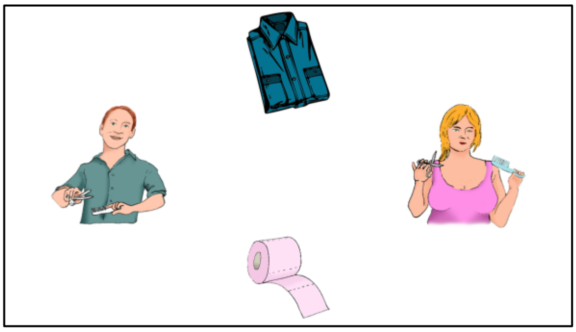
F25



F26

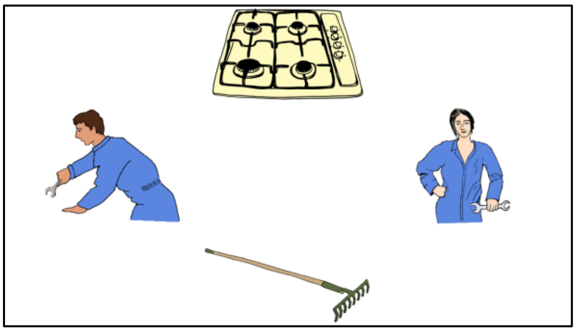


F27

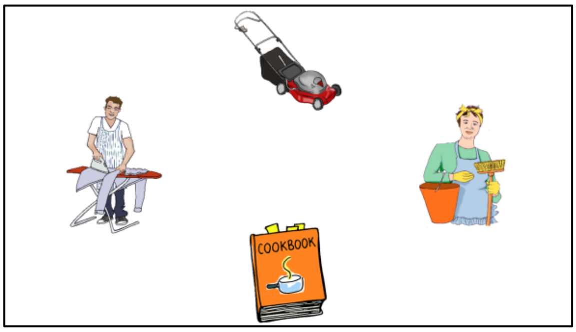


F28

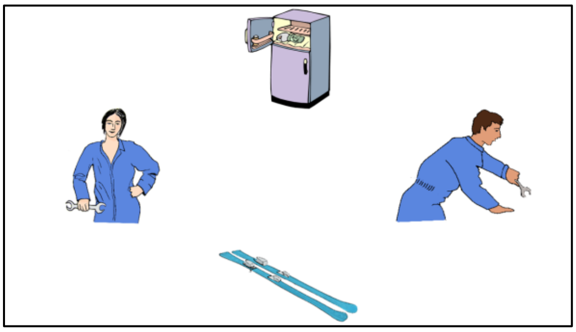
Each of the visual scenes below represents the content of a filler sentence. The number of the filler sentence a visual scene refers to is indicated below (see Appendix A, Table A18 for the full set of filler sentences) (continued).



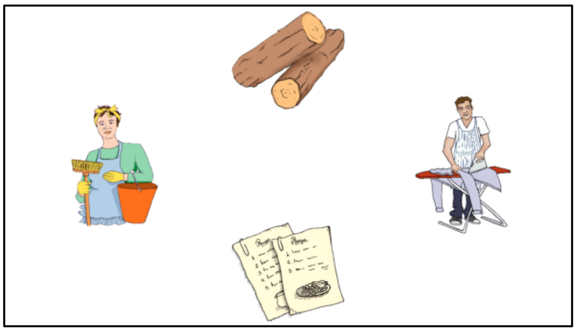
F29



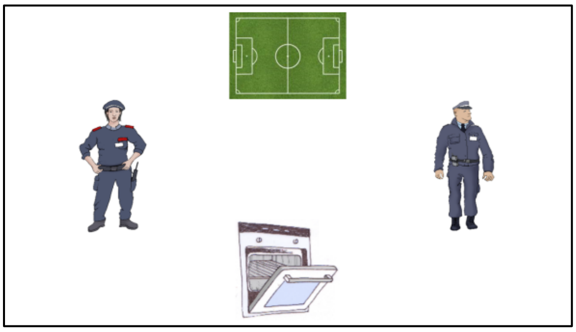
F30



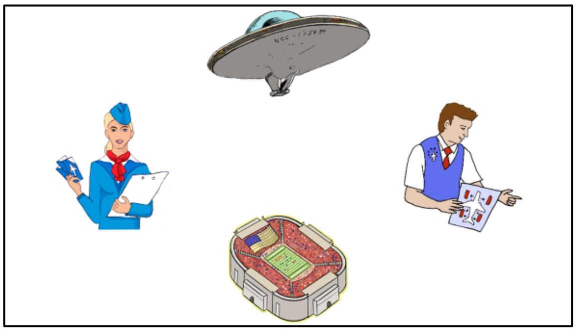
F31



F32

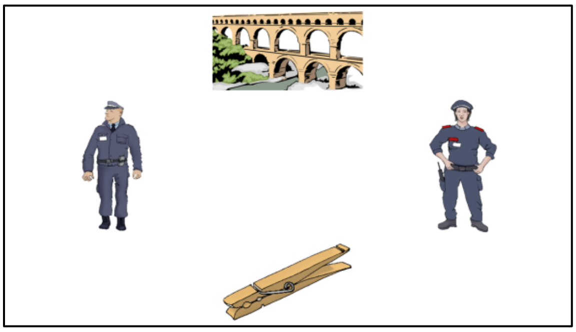


F33

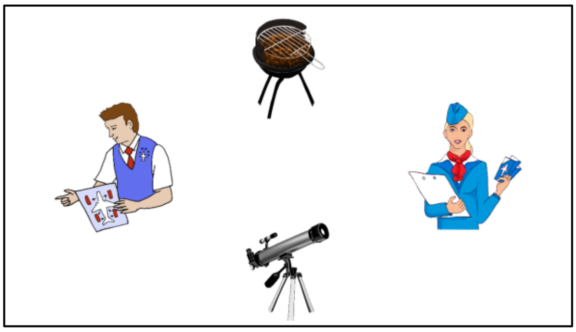


F34

Each of the visual scenes below represents the content of a filler sentence. The number of the filler sentence a visual scene refers to is indicated below (see Appendix A, Table A18 for the full set of filler sentences) (continued).



F35

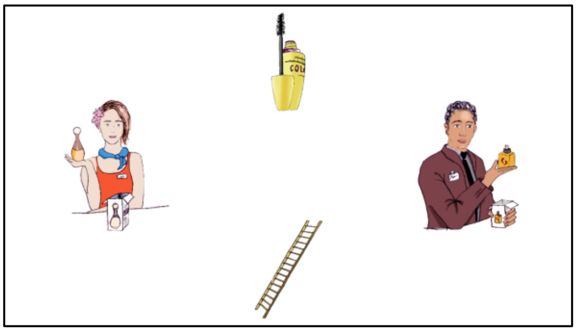


F36

F37

F38

F39



F40

A Experiment 1 (Adverbs – Humans) – Questionnaire: The Measured Constructs

All items that were used in the questionnaire are listed in the following.

Table A19

Items of the Ambivalent Sexism Inventory (ASI; Glick & Fiske. 1996; German Version: Eckes & Six-Materna, 1999)

No.	Statement
1	Egal wie erfolgreich ein Mann auch sein mag, ohne eine Frau, die ihn liebt, fehlt ihm etwas ganz Wichtiges. [No matter how accomplished he is, a man is not truly complete as a person unless he has the love of a woman.]
2	Viele Frauen versuchen unter dem Deckmantel der Gleichberechtigung besondere Vergünstigungen zu erlangen wie z.B. eine Bevorzugung bei der Besetzung von Arbeitsstellen. [Many women are actually seeking special favors, such as hiring policies that favor them over men, under the guise of asking for "equality".]
3	Bei einer Katastrophe sollten Frauen vor Männern gerettet werden. [In a disaster, women ought to be rescued before men.]
4	Die meisten Frauen interpretieren harmlose Äußerungen oder Handlungen als frauenfeindlich. [Most women interpret innocent remarks or acts as being sexist.]
5	Frauen sind zu schnell beleidigt. [Women are too easily offended.]
6	Ein Mann kann im Leben erst richtig glücklich sein, wenn er eine Partnerin hat, die er liebt. [A man cannot be truly happy in life without a woman he loves.]
7	Was Feministinnen wirklich wollen ist, dass Frauen mehr Macht bekommen als Männer. [Feminists are seeking for women to have more power than men.]
8	Viele Frauen haben eine Art von Ehrlichkeit, die nur wenige Männer besitzen. [Many women have a quality of purity that few men possess.]
9	Frauen sollten von Männern umsorgt und beschützt werden. [Women should be cherished and protected by men.]
10	Die meisten Frauen sehen gar nicht, was Männer alles für sie tun. [Most women fail to appreciate fully all what men do for them.]
11	Frauen versuchen Macht zu erlangen, indem sie Männer immer mehr beherrschen. [Women seek to gain power by getting control over men.]

Note. 7-point Likert scales indicating 1 = I strongly 7 = I agree strongly. Items on benevolent sexism: 1, 3, 6, 8, 9, 12, 13, 17, 19, 20, 22, item on hostile sexism: 2, 4, 5, 7, 10, 11, 14, 15, 16, 18, 21. The item's translation to English is inserted in parentheses below.

(continued)

Table A19

Items of the Ambivalent Sexism Inventory (ASI; Glick & Fiske. 1996; German Version: Eckes & Six-Materna, 1999) (continued)

No.	Statement
12	Jeder Mann sollte eine Frau haben, die er richtig liebt. [Every man ought to have a woman whom he adores.]
13	Männer sind ohne Frauen unvollkommen. [Men are incomplete without women.]
14	Frauen übertreiben Probleme, die sie am Arbeitsplatz haben. [Women exaggerate problems they have at work.]
15	Hat eine Frau erst einmal einen Mann „rumgekriegt", dann versucht sie, ihn an die kurze Leine zu legen. [Once a woman gets a man to commit to her, she usually tries to put him on a tight leash.]
16	Wenn Frauen in einem fairen Wettbewerb gegenüber Männern den Kürzeren ziehen, behaupten sie gerne, sie seien diskriminiert worden. [When women lose to men in a fair competition, they typically complain about being discriminated against.]
17	Eine Frau sollte von ihrem Mann auf Händen getragen werden. [A good woman should be set on a pedestal by her man.]
18	Viele Frauen haben Spaß daran, mit Männern zu „spielen", indem sie sich zuerst verführerisch geben, dann aber die Annäherungsversuche der Männer zurückweisen. [Many women get a kick out of teasing men by seeming sexually available and then refusing male advances.]
19	Verglichen mit Männern haben Frauen ein besseres moralisches Empfinden. [Women, compared to men, tend to have a superior moral sensibility.]
20	Ein Mann sollte bereit sein, sein eigenes Wohl zu opfern, um für seine Frau sorgen zu können. [Men should be willing to sacrifice their own well-being in order to provide financially for the women of their lives.]
21	Feministinnen stellen an Männer vollkommen überzogene Forderungen. [Feminists are making entirely excessive demands of men.]
22	Verglichen mit Männern haben Frauen einen feineren Sinn für Kultur und einen besseren Geschmack. [Women, as compared to men, tend to have a more refined sense of culture and good taste.]

Note. 7-point Likert scales indicating 1 = I totally refuse 7 = I totally agree. Items on benevolent sexism: 1, 3, 6, 8, 9, 12, 13, 17, 19, 20, 22, items on hostile sexism: 2, 4, 5, 7, 10, 11, 14, 15, 16, 18, 21. The item's translation to English is inserted in parentheses below.

Table A20

Items of the Normative Gender Role Orientation Scale (NGRO; Athenstaedt, 2000)

No.	Statement
1	Auch Männer sollten nach der Geburt ihres Kindes die Möglichkeit einer Elternzeit in Anspruch nehmen können. [Men should also be able to go on parental leave after the birth of their child.]
2	Es ist angenehmer, einen männlichen Vorgesetzten zu haben als einen weiblichen. [It is more pleasant to have a male superior than a female one.]
3	Jungen und Mädchen sollen die gleichen Pflichten im Haushalt übernehmen. [Boys and girls should have the same household obligations.]
4	Frauen sind weniger an Politik interessiert als Männer. [Women are less interested in politics than men.]
5	Man kann von Frauen nicht fordern, dass sie die Hausarbeit alleine verrichten müssen. [Women cannot be required to do all the domestic work alone.]
6	Für den ersten Eindruck ist ein gepflegtes Äußeres bei einer Frau wichtiger als bei einem Mann. [For the first impression, a groomed appearance is more important for women than for men.]
7	Auch der Mann hat dafür zu sorgen, dass täglich Milch und Brot im Haus sind. [Also, the man should care about having milk and bread at home.]
8	Frauen lassen sich gerne von ihrem männlichen Begleiter einladen. [Women enjoy being invited by their male companion.]
9	Hemden zu bügeln ist nicht Sache der Männer. [To iron shirts is not men's business.]
10	Eine höhere Ausbildung ist vor allem für Männer wichtig, da sie in Führungspositionen stärker vertreten sind als Frauen. [Higher education is especially important for men because they are more strongly represented in leadership positions than women.]
11	Frauen eignen sich ebenso gut für die Leitung eines technischen Betriebes wie Männer. [Women are equally suitable to manage a technical running as men.]
12	Männer sollten in der Politik mehr auf Frauen hören. [In politics, men should take women's advice more often.]
13	Es wäre erfreulich. wenn es in KiTas mehr männliche Erzieher gäbe. [It would be nice having more male educators in day care centers for children.]
14	Männer sind für manche Berufe besser geeignet als Frauen. [Men are more suitable for some jobs than women.]
15	Jeder Junge sollte eine Puppe besitzen. [Every boy should own a doll.]

Note. 7-point Likert scales indicating 1 = I totally refuse 7 = I totally agree. Egalitarian items had to be reversed. Reversed items: 1, 3, 5, 7, 11, 12, 13, 15, 17, 18, 21, 26, 28, 29. The item's translation to English is inserted in parentheses below.

(continued)

Table A20

Items of the Normative Gender Role Orientation Scale (NGRO; Athenstaedt, 2000) (continued)

No.	Statement
16	Mädchen helfen lieber im Haushalt als Jungen. [Girls prefer helping with housework rather than boys.]
17	Die Putztätigkeit sollte auf beide Ehepartner entsprechend ihrer verfügbaren Zeit. aufgeteilt werden. [Cleaning work should be split between spouses according to their available time.]
18	Der Anteil der Frauen in der Politik sollte gleich groß sein wie der Anteil der Männer. [The proportion of women and men should be equal in politics.]
19	Das Vertrauen in Politikerinnen ist nicht so groß, da diese meistens noch andere Dinge als ihr Amt im Kopf haben. [People have less trust in female politicians because they commonly have other things in mind apart from their political duties.]
20	Dass Männer im Allgemeinen mehr verdienen, liegt daran, dass sie sich beruflich mehr einsetzen als Frauen. [That men commonly earn more money is because they focus more strongly on their professional career than women.]
21	Eine Frau kann das Verteidigungsministerium ebenso gut leiten wie ein Mann. [A woman can lead the Department of Defense as good as a man can do.]
22	Männliche Polizisten vermitteln ein stärkeres Sicherheitsgefühl als Polizistinnen. [Male police officers provide a stronger sense of security than female police officers.]
23	Die Organisation des Haushaltes ist Sache der Frau. [Organizing domestic chores is the duty of a woman.]
24	Es ist notwendig, dass die Frau im Hause dafür sorgt, dass täglich zumindest eine warme Mahlzeit auf dem Tisch steht. [It is necessary that the woman provides at least one warm meal a day.]
25	Es ist nicht in Ordnung, wenn eine Frau den Rasen mäht, während ihr Mann das Essen kocht. [It is inappropriate for a woman to mow the lawn while her husband prepares the meal.]
26	Auch Hausmann ist für Männer ein erstrebenswerter Beruf. [Being a househusband is a desirable profession also for men.]
27	Meistens haben Frauen die größere Verantwortung für den Haushalt, weil sie ihn besser führen können. [Women commonly have greater responsibilities for household maintenance because they are more suitable of caring for.]
28	Männer sollten sich auch mit Handarbeit (z.B. Nähen, Stricken) beschäftigen. [Men should also care about needlework (e.g., sewing, knitting)].
29	Frauen sind für den finanziellen Unterhalt der Familie genauso verantwortlich wie Männer. [Women are equally responsible for the financial maintenance of the family as men.]

Note. 7-point Likert scales indicating 1 = I totally refuse 7 = I totally agree. Egalitarian items had to be reversed. Reversed items: 1, 3, 5, 7, 11, 12, 13, 15, 17, 18, 21, 26, 28, 29. The item's translation to English is inserted in parentheses below.

Table A21

Items of the Social Desirability Scale (SDS-17; Stöber, 2001; German Version: Stöber, 1999)

No.	Statement
1	Manchmal werfe ich Müll einfach in die Landschaft oder auf die Straße. [I sometimes litter.]
2	Eigene Fehler gebe ich stets offen zu und ertrage gelassen etwaige negative Konsequenzen. [I always admit my mistakes openly and face the potential negative consequences.]
3	Im Straßenverkehr nehme ich stets Rücksicht auf die anderen Verkehrsteilnehmer. [In traffic I am always polite and considerate of others.]
4	Ich habe schon einmal Drogen (Tabletten, Haschisch oder "ähnliches") konsumiert. [I have tried illegal drugs (for example, marijuana, cocaine, etc.).]
5	Ich akzeptiere alle anderen Meinungen, auch wenn sie mit meiner eigenen nicht übereinstimmen. [I always accept others' opinions, even when they don't agree with my own.]
6	Meine Wut oder schlechte Laune lasse ich hin und wieder an unschuldigen oder schwächeren Leuten aus. [I take out my bad moods on others now and then.]
7	Ich habe schon einmal jemanden ausgenutzt oder übers Ohr gehauen. [There has been an occasion when I took advantage of someone else.]
8	In einem Gespräch lasse ich den anderen stets ausreden und höre ihm aufmerksam zu. [In conversations I always listen attentively and let others finish their sentences.]
9	Ich zögere niemals, jemanden in einer Notlage beizustehen. [I never hesitate to help someone in case of emergency.]
10	Wenn ich etwas versprochen habe, halte ich es ohne Wenn und Aber. [When I have made a promise, I keep it – no ifs, ands, or buts.]
11	Ich lästere gelegentlich über andere hinter deren Rücken. [I occasionally speak badly of others behind their back.]
12	Ich würde niemals auf Kosten der Allgemeinheit leben. [I would never live off other people.]
13	Ich bleibe immer freundlich und zuvorkommend anderen Leuten gegenüber, auch wenn ich selbst gestresst bin. [I always stay friendly and courteous with other people, even when I am stressed out.]
14	Im Streit bleibe ich stets sachlich und objektiv. [During arguments I always stay objective and matter-of-fact.]
15	Ich habe schon einmal geliehene Sachen nicht zurückgegeben. [There has been at least one occasion when I failed to return an item that I borrowed.]
16	Ich ernähre mich stets gesund. [I always eat a healthy diet.]
17	Manchmal helfe ich nur, weil ich eine Gegenleistung erwarte. [Sometimes I only help because I expect something in return.]

Note. 7-point Likert scales indicating 1 = I totally refuse 7 = I totally agree. Reversed items: 1, 4, 6, 7, 11, 15, 17. The English item (Stöber, 2001) is inserted in parentheses below.

Table A22

Items of the Motivation to Control for Sexist Responses Scale (MCSR; Klonis et al.,2005; German Version: Eyssel, 2010)

No.	Statement
1	Meinen persönlichen Werten zufolge ist es in Ordnung, sich Stereotypen gegenüber Frauen zu bedienen. [According to my personal values, using stereotypes about women is OK.]
2	Ich bin persönlich motiviert durch meine Überzeugungen nichtsexistisch gegenüber Frauen zu sein. [I am personally motivated by my beliefs to be nonsexist toward women.]
3	Nichtsexistisch gegenüber Frauen zu sein ist wichtig für mein Selbstkonzept. [Being nonsexist toward women is important to my self-concept.]
4	Aufgrund meiner persönlichen Werte/Werthaltungen glaube ich, dass es falsch ist, sich Stereotype über Frauen zu bedienen. [Because of my personal values, I believe that using stereotypes about women is wrong.]
5	Ich versuche mich nichtsexistisch gegenüber Frauen zu verhalten, weil es mir persönlich wichtig ist. [I attempt to act in nonsexist ways toward women because it is personally important to me.]
6	Ich unterstütze gleiche Rechte für Frauen, weil es mir persönlich wichtig ist. [I support equal rights for women because it is personally important to me.]
7	Aufgrund meiner persönlichen Überzeugungen denke ich, dass Frauen ohne Widerstand der Gesellschaft männerdominierte Berufe ausüben können sollten. [Because of my personal beliefs, I think women should be able to go into male-dominated careers without resistance from society.]
8	Meinen persönlichen Überzeugungen zufolge sollten Frauen genauso viel sexuelle Freiheit haben wie Männer. [According to my personal beliefs, women should have as much sexual freedom as men have.]
9	Es ist persönlich wichtig für mich, Leute wissen zu lassen, dass ich denke Frauen sind genauso gut wie Männer in einer höheren Berufslaufbahn. [It is personally important to me to let people know that I think women are just as good as men in high-level careers.]
10	Meinen persönlichen Standards zufolge, steht Frauen genauso wie Männern Zugang zu Führungspositionen zu. [According to my personal standards, women are entitled to have as much access to leadership roles as men.]

Note. 7-point Likert scales indicating 1 = I totally refuse 7 = I totally agree. Items on participants' intrinsic motivation: 1 – 10, items on participants' extrinsic motivation: 11 – 20. Item 1 had to be reversed. The English item (Klonis et al., 2005) is inserted in parentheses below.

(continued)

Table A22

Items of the Motivation to Control for Sexist Responses Scale (MCSR; Klonis et al.,2005; German Version: Eyssel, 2010) (continued)

No.	Statement
11	Aufgrund der heutigen Standards politischer Korrektheit, versuche ich nichtsexistisch gegenüber Frauen zu erscheinen. [Because of today's PC (politically correct) standards I try to appear nonsexist toward women.]
12	Ich versuche irgendwelche negativen Gedanken über Frauen zu verbergen, um negative Reaktionen von anderen zu vermeiden. [try to hide any negative thoughts about women in order to avoid negative reactions from others.]
13	Wenn ich mich sexistisch gegenüber Frauen verhalten würde, wäre ich beunruhigt, dass andere ärgerlich auf mich sein würden. [If I acted sexist toward women, I would be concerned that others would be angry with me.]
14	Ich versuche, nichtsexistisch gegenüber Frauen zu erscheinen, um eine Missbilligung von anderen zu vermeiden. [I attempt to appear nonsexist toward women in order to avoid disapproval from others.]
15	Aufgrund von Druck von anderen versuche ich, mich nichtsexistisch zu verhalten. [I try to act in nonsexist ways because of pressure from others.]
16	Ich unterstütze Frauenrechte, weil ich fühle, dass ich es im heutigen Klima politischer Korrektheit muss. [I support women's rights because I feel like I have to, in today's PC climate.]
17	Ich versuche, Frauen und Männer als ebenbürtig zu behandeln, weil ich Angst habe, dass andere verärgert über mich wären, wenn ich es nicht täte. [I try to treat women and men as equals, because I'm afraid other people would be upset with me if I didn't.]
18	Ich unterstütze Frauen, die in männer-dominierten Berufe gehen öffentlich, weil ich Angst vor der Missbilligung anderer habe. [I publicly support women going into male-dominated careers, because I'm afraid of disapproval from others.]
19	Wegen sozialem Druck versuche ich, mir nicht anmerken zu lassen, dass ich denke, dass Frauen besser in ihre traditionelle Rolle passen. [Because of social pressure, I try not to let on that I think women are better-suited for their traditional roles.]
20	Ich erzähle keine Blondinenwitze vor Leuten, von denen ich denke, dass sie mich Sexist nennen könnten. [I don't tell blonde jokes around people who I think might call me a sexist.]

Note. 7-point Likert scales indicating 1 = I totally refuse 7 = I totally agree. Items on participants' intrinsic motivation: 1 – 10, items on participants' extrinsic motivation: 11 – 20. Item 1 had to be reversed. The English item (Klonis et al., 2005) is inserted in parentheses below.

A Experiment 1 (Adverbs – Humans) – Experiment Instructions

The German wording of the experiment instructions (Guerra et al., 2021, adapted) that were displayed on the screen (see Section 4.10) and its translation to English in a paragraph below:

Welcoming participants: “Willkommen und danke für dein Interesse an dieser Studie. Zuerst bitten wir dich, die Informationen zur Studie zu lesen. Drücke bitte den blauen Knopf.“

[Welcome and thank you for participating in this study. First, we ask you to carefully read the information on the study. Please, press the blue button.]

> **German instructions, Part 1:** Liebe Teilnehmerin, lieber Teilnehmer, in dieser Studie werden dir verschiedene Bilder präsentiert. Schau dir diese bitte genau an. Kurz nachdem das Bild eingeblendet wurde, wirst du einen Satz hören, der sich auf das eingeblendete Bild bezieht. Bitte achte ebenfalls aufmerksam auf den Inhalt des Satzes.
>
> Gelegentlich werden auf dem Bildschirm Fragen eingeblendet. Beantworte diese bitte mit „JA“ (durch Betätigen der JA-Taste) oder mit „NEIN“ (durch Betätigen der NEIN-Taste). Die Fragen können sich auf den Inhalt der Bilder oder der Sätze beziehen. Daher ist es wichtig, dass du aufmerksam auf die Bilder und die Sätze achtest. Falls du hierzu Fragen hast, frag’ ruhig nach!

[Dear participant, in this study, various pictures will be presented on the screen. Please look at them carefully. Shortly after the picture’s onset, you will hear a sentence referring to the picture. Please pay close attention to the sentences as well. Some trials will be followed by a question. Please answer these questions by indicating “yes” (by pressing the yes-button) or “no” (by pressing the no-button). The questions concern either the content of the pictures or the sentences. It is thus very important that you pay close attention to the pictures as to the sentences. If you have any questions, don’t hesitate to ask.]

> **German instructions, Part 2:** Es ist sehr wichtig, dass du während der gesamten Zeit auf den Bildschirm schaust. Des Weiteren solltest du dich auf das Verstehen der Sätze konzentrieren, sodass du dir sicher sein kannst, die richtige Antwort zu geben. Wenn du die Erklärungen verstanden hast, drücke bitte den

blauen Knopf, um zu beginnen. Wir wünschen dir jetzt viel Erfolg und Spaß bei der Studie.[It is important that you keep your eyes on the monitor all the time. Additionally, please listen to the sentences carefully to make sure to answer the questions correctly. If you understood the instructions, please press the blue button to start. Good luck and have fun participating in the study.]

End: “Danke für deine Teilnahme.” **[**Thanks for your participation.]

A Pretest IV – Results: One-Sample *t*-Tests on Noun and Verb Phrases' Gender-Stereotypicality

Table A23

Noun Phrases Considered Stereotypically Male According to a One-Sample t-Test Against the Scale Midpoint of 4 and the Verb Phrases That Followed the Noun Phrases in Pretest IV

Noun Phrase	Following Verb Phrase	*M*	*SD*	*t*	*df*	*p*	Cohen's *d*
das Motorboot [the motorboat]	wird gesteuert [is steered]	1.95	0.92	-13.98	38	< .001	2.24
das Pokerspiel [the poker game]	wird verfolgt [is followed]	2.08	1.04	-11.60	38	< .001	1.86
das Motorboot [the motorboat]	wird zerkratzt [is scratched]	2.18	1.10	-1.36	38	< .001	1.66
das Motorboot [the motorboat]	wird betreten [is entered]	2.21	1.11	-1.15	38	< .001	1.63
das Pokerspiel [the poker game]	wird verachtet [is despised]	2.23	1.04	-1.65	38	< .001	1.71
das Pokerspiel [the poker game]	wird erklärt [is explained]	2.38	1.37	-7.37	38	< .001	1.18
das Raumschiff [the spaceship]	wird ekundet [is explored]	2.41	1.12	-8.89	38	< .001	1.42
das Raumschiff [the spaceship]	wird gesichtet [is sighted]	2.46	1.14	-8.40	38	< .001	1.35
das Raumschiff [the spaceship]	wird gemieden [is avoided]	2.49	1.12	-8.43	38	< .001	1.36
das Holz [the wood]	wird gehackt [is chopped]	2.64	1.11	-7.63	38	< .001	1.22
das Holz [the wood]	wird angemalt [is painted]	2.67	1.13	-7.36	38	< .001	1.18
das Holz [the wood]	wird gelagert [is stored]	2.77	1.18	-6.51	38	< .001	1.04
das Teleskop [the telescope]	wird verwahrt [is kept safe]	2.85	0.61	-7.50	38	< .001	1.20
das Teleskop [the telescope]	wird aufgestellt [positioned]	2.87	0.95	-7.41	38	< .001	1.19
das Teleskop [the telescope]	wird umgestoßen [is knocked over]	3.23	0.96	-5.01	38	< .001	0.80
das Ladegerät [the charger]	wird angeschlossen [is connected]	3.28	1.00	-4.49	38	< .001	0.72
das Ladegerät [the charger]	wird aufgerollt [is rolled]	3.33	0.98	-4.24	38	< .001	0.68
das Tier [the animal]	wird geschossen [is shot]	3.51	1.14	-2.66	38	.011	0.43
das Brot [the bread]	wird gebacken [is baked]	3.54	0.97	-2.97	38	.005	0.48
das Tablet [the tablet]	wird gekauft [is bought]	3.54	1.21	-2.38	38	.022	0.38
das Fachbuch [the reference book]	wird eingesteckt [is packed]	3.54	0.79	-3.65	38	.001	0.58
das Handy [the cell phone]	wird aufbewahrt [is stored]	3.56	0.91	-2.99	38	.005	0.48
das Opossum [the opossum]	wird angelächelt [is smiled]	3.56	1.14	-2.38	38	.022	0.38
das Tier [the animal]	wird gestreichelt [is stroked]	3.59	1.07	-2.40	38	.022	0.38

Note. 1 = stereotypically male, 7 = stereotypically female. The noun phrase's (marked in bold) and the following verb phrase's translation to English are inserted in parentheses below.

(continued)

Table A23

Noun Phrases Considered Stereotypically Male According to a One-Sample t-Test Against the Scale Midpoint of 4 and the Verb Phrases That Followed the Noun Phrases in Pretest IV (continued)

Noun Phrase	Following Verb Phrase	*M*	*SD*	*t*	*df*	*p*	Cohen's *d*
das Tablet [the tablet]	wird upgedated [is updated]	3.59	1.21	-2.12	38	.040	0.34
das Labyrinth [the labyrinth]	wird gezeichnet [is drawn]	3.62	0.91	-2.65	38	.012	0.42
das Dokument [the document]	wird vorgelesen [is read]	3.62	0.82	-2.95	38	.005	0.47
das Opossum [the opossum]	wird verspottet [is mocked]	3.62	0.82	-2.95	38	.005	0.47
das Regal [the shelf]	wird justiert [is adjusted]	3.64	0.96	-2.34	38	.025	0.37
das Regal [the shelf]	wird entstaubt [is dusted]	3.67	0.77	-2.70	38	.010	0.43
das Haus [the house]	wird errichtet [is built]	3.69	0.92	-2.08	38	.044	0.33
das Regal [the shelf]	wird dekoriert [is decorated]	3.69	0.92	-2.08	38	.044	0.33
das Regal [the shelf]	wird montiert [is mounted]	3.69	0.77	-2.51	38	.017	0.40
das Radio [the radio]	wird bewundert [is admired]	3.72	0.69	-2.57	38	.014	0.41
das Regal [the shelf]	wird ausgewählt [is chosen]	3.74	0.72	-2.24	38	.031	0.40
das Labyrinth [the labyrinth]	wird vermessen [measured]	3.74	0.79	-2.04	38	.048	0.33

Note. 1 = stereotypically male, 7 = stereotypically female. The noun phrase's (marked in bold) and the following verb phrase's translation to English are inserted in parentheses below.

Table A24

Noun Phrases Considered Stereotypically Female According to a One-Sample t-Test Against the Scale Midpoint of 4 and the Verb Phrases That Followed the Noun Phrases in Pretest IV

Noun Phrase	Following Verb Phrase	*M*	*SD*	*t*	*df*	*p*	Cohen's *d*
das Kind [the child]	wird bemuttert [is mothered]	4.23	0.67	2.16	38	.037	0.35
das Kind [the child]	wird bespaßt [is entertained]	4.31	0.80	2.40	38	.021	0.39
das Kind [the child]	wird beschenkt [is made a present]	4.36	0.84	2.66	38	.011	0.43
das Klavier [the piano]	wird restauriert [is restored]	4.49	1.02	2.98	38	.005	0.48
das Gift [the poison]	wird erforscht [is researched]	4.49	1.49	2.05	38	.047	0.33
das Kuscheltier [the cuddly toy]	wird verbrant [is burnt]	4.77	0.93	5.16	38	< .001	0.83
das Kuscheltier [the cuddly toy]	wird gegeben [is given]	4.82	1.05	4.89	38	< .001	0.78
das Kuscheltier [the cuddly toy]	wird umarmt [is hugged]	4.92	1.09	5.31	38	< .001	0.85
das Pferd [the horse]	wird gestriegelt [is curried]	5.23	1.42	5.40	38	< .001	0.87
das Pferd [the horse]	wird geritten [is ridden]	5.26	1.37	5.72	38	< .001	0.92
das Pferd [the horse]	wird verarztet [is doctored]	5.26	1.50	5.23	38	< .001	0.84
das Kleid [the dress]	wird angezogen [is put on]	6.41	1.09	13.77	38	< .001	2.20
das Kleid [the dress]	wird gestohlen [is stolen]	6.49	1.02	15.19	38	< .001	2.43
das Kleid [the dress]	wird komplimentiert [is complimented]	6.51	0.91	17.17	38	< .001	2.75

Note. 1 = stereotypically male, 7 = stereotypically female. The noun phrase's (marked in bold) and the following verb phrase's translation to English are inserted in parentheses below.

Table A25

Noun Phrases Considered Stereotypically Gender-Neutral According to a One-Sample t-Test Against the Scale Midpoint of 4 and the Verb Phrases That Followed the Noun Phrases in Pretest IV

Noun Phrase	Following Verb Phrase	*M*	*SD*	*t*	*df*	*p*	Cohen's *d*
das Fischfilet [the fish fillet]	wird abgetaut [is defrosted]	3.69	1.03	-1.87	38	.070	0.30
das Radio [the radio]	wird reguliert [is regulated]	3.72	0.89	-1.99	38	.054	0.32
das Radiergummi [the eraser]	wird eingepackt [is packed]	3.74	0.97	-1.66	38	.106	0.27
das Ei [the egg]	wird aufgeschlagen [is cracked]	3.74	1.04	-1.53	38	.133	0.25
das Aquarium [the aquarium]	wird aufgestellt [is positioned]	3.74	1.27	-1.26	38	.216	0.20
das Sofa [the sofa]	wird ausgesucht [is chosen]	3.77	0.78	-1.86	38	.071	0.30
das Brot [the bread]	wird gebrochen [is broken]	3.77	0.96	-1.50	38	.141	0.24
das Sofa [the sofa]	wird verladen [is loaded]	3.77	0.96	-1.50	38	.141	0.24
das Handy [the cell phone]	wird repariert [is repaired]	3.79	0.77	-1.67	38	.103	0.27
das Sofa [the sofa]	bepolstert [is cushioned]	3.79	0.77	-1.67	38	.103	0.27
das Spielzeug [the toy]	wird desinfiziert [is disinfected]	3.79	0.80	-1.60	38	.118	0.26
das Dokument [the document]	wird gelöscht [is deleted]	3.79	0.83	-1.54	38	.132	0.25
das Sofa [the sofa]	wird genäht [is sewn]	3.79	0.86	-1.48	38	.146	0.24
das Paket [the parcel]	wird angehoben [is lifted]	3.79	0.89	-1.43	38	.160	0.23
das Blatt [the sheet]	wird entworfen [is designed]	3.82	0.64	-1.74	38	.090	0.28
das Sofa [the sofa]	wird dekoriert [is decorated]	3.82	0.79	-1.42	38	.164	0.23
das Naturereignis [the natural event]	wird prognostiziert [is forecasted]	3.82	1.00	-1.13	38	.268	0.18
das Fachbuch [the reference book]	wird gelesen [is read]	3.82	1.05	-1.07	38	.292	0.17
das Ei [the egg]	wird gekocht [is boiled]	3.82	1.30	-0.87	38	.392	0.14
das Zimmer [the room]	wird eingerichtet [is furnished]	3.85	0.67	-1.43	38	.160	0.23
das Plakat [the poster]	wird ausgeschmückt [is embellished]	3.85	0.88	-1.10	38	.279	0.18
das Naturereignis [the natural event]	wird gefürchtet [is feared]	3.85	0.90	-1.06	38	.295	0.17
das Album [the album]	wird erkundet [is explored]	3.87	0.86	-0.93	38	.360	0.15
das Brettspiel [the board game]	wird eingebaut [is installed]	3.87	0.98	-0.82	38	.418	0.13
das Radiergummi [the eraser]	wird angefertigt [is manufactured]	3.87	1.15	-0.70	38	.491	0.11
das Brettspiel [the board game]	wird organisiert [is organized]	3.90	0.85	-0.75	38	.457	0.12

Note. 1 = stereotypically male, 7 = stereotypically female. The noun phrase's (marked in bold) and the following verb phrase's translation to English are inserted in parentheses below.

(continued)

Table A25

Noun Phrases Considered Stereotypically Gender-Neutral According to a One-Sample t-Test Against the Scale Midpoint of 4 and the Verb Phrases That Followed the Noun Phrases in Pretest IV (continued)

Noun Phrase	Following Verb Phrase	*M*	*SD*	*t*	*df*	*p*	Cohen's *d*
das Aquarium [the aquarium]	wird gereinigt [is cleaned]	3.90	0.88	-0.26	38	.472	0.04
das Wasser [the water]	wird gefärbt [is colored]	3.92	0.62	-0.77	38	.446	0.12
das Publikum [the audience]	wird evakuiert [is evacuated]	3.92	0.70	-0.68	38	.498	0.11
das Brötchen [the roll]	wird aufgeschnitten [is sliced open]	3.92	0.87	-0.55	38	.584	0.09
das Spielzeug [the toy]	wird gesteuert [is steered]	3.92	0.93	-0.52	38	.608	0.08
das Getränk [the drink]	wird gezapft [is tapped]	3.92	0.98	-0.49	38	.628	0.08
das Fischfilet [the fish fillet]	wird gegrillt [is roasted]	3.92	1.31	-0.37	38	.715	0.06
das Blatt [the sheet]	wird zerstört [is destroyed]	3.95	0.69	-0.47	38	.643	0.08
das Haus [the house]	wird aufgeräumt [is tidied]	3.95	1.12	-0.29	38	.777	0.05
das Paket [the parcel]	wird bestellt [is ordered]	3.95	1.15	-0.28	38	.781	0.05
das Wasser [the water]	wird getestet [is tested]	3.97	0.78	-0.21	38	.838	0.03
das Heft [the booklet]	wird ausgefüllt [is filled]	3.97	0.81	-0.20	38	.844	0.03
das Brötchen [the roll]	wird zubereitet [is prepared]	3.97	0.96	-0.17	38	.868	0.03
das Mobiliar [the furniture]	wird aufgebaut [is built up]	3.97	1.01	-0.16	38	.875	0.03
das Eichhörnchen [the squirrel]	wird verfolgt [is followed]	3.97	1.11	-0.14	38	.886	0.02
das Publikum [the audience]	wird wahrgenommen [is perceived]	4.00	0.86	< 0.01	38	1.00	< 0.01
das Museum [the museum]	wird geräumt [is evacuated]	4.00	0.86	< 0.01	38	1.00	< 0.01
das Glücksrad [the prize wheel]	wird lackiert [is varnished]	4.00	1.08	< 0.01	38	1.00	< 0.01
das Fenster [the window]	wird ausgemessen [is sized]	4.03	0.71	0.23	38	.822	0.04
das Handtuch [the towel]	wird gebügelt [is ironed]	4.03	0.84	0.19	38	.850	0.03
das Plakat [the poster]	wird verschönert [is beautified]	4.08	1.09	0.44	38	.661	0.07
das Museum [the museum]	wird besucht [is visited]	4.08	1.09	0.44	38	.661	0.07
das Eichhörnchen [the squirrel]	wird gewiegt [is cradled]	4.08	1.11	0.43	38	.667	0.07
das Zimmer [the room]	wird gestrichen [is painted]	4.08	1.13	0.42	38	.674	0.07
das Xylophon [the xylophone]	wird angehört [is listened]	4.10	1.35	0.47	38	.639	0.08
das Album [the album]	wird bemalt [is painted]	4.13	0.95	0.84	38	.405	0.14

Note. 1 = stereotypically male, 7 = stereotypically female. The noun phrase's (marked in bold) and the following verb phrase's translation to English are inserted in parentheses below.

(continued)

Table A25

Noun Phrases Considered Stereotypically Gender-Neutral According to a One-Sample t-Test Against the Scale Midpoint of 4 and the Verb Phrases That Followed the Noun Phrases in Pretest IV (continued)

Noun Phrase	Following Verb Phrase	*M*	*SD*	*t*	*df*	*p*	Cohen's *d*
das Fenster [the window]	wird geputzt [is cleaned]	4.13	0.95	0.84	38	.405	0.14
das Glücksrad [the prize wheel]	wird präsentiert [is presented]	4.13	1.11	0.73	38	.473	0.12
das Portemonnaie [the wallet]	wird gefunden [is found]	4.13	1.34	0.60	38	.554	0.10
das Getränk [the drink]	wird umgerührt [is stirred up]	4.15	0.59	1.64	38	.110	0.26
das Heft [the booklet]	wird gedruckt [is printed]	4.15	0.67	1.43	38	.160	0.23
das Handtuch [the towel]	wird beschmutzt [is dirtied]	4.21	1.17	1.09	38	.282	0.18
das Xylophon [the xylophone]	wird gestimmt [is tuned]	4.21	1.24	1.03	38	.308	0.17
das Mobiliar [the furniture]	wird ausgewählt [is chosen]	4.26	1.02	1.57	38	.124	0.25
das Klavier [the piano]	wird ausgesucht [is chosen]	4.28	1.03	1.72	38	.094	0.28
das Gift [the poison]	wird versprüht [is sprayed]	4.31	1.10	1.74	38	.090	0.28
das Portemonnaie [the wallet]	wird produziert [is produced]	4.31	1.22	1.58	38	.123	0.25

Note. 1 = stereotypically male, 7 = stereotypically female. The noun phrase’s (marked in bold) and the following verb phrase’s translation to English are inserted in parentheses below.

Table A26

Verb Phrases Considered Stereotypically Male According to a One-Sample t-Test Against the Scale Midpoint of 4 and the Noun Phrases That Anteceded the Verb Phrases in Pretest IV

Anteceding Noun Phrase	Verb Phrase	*M*	*SD*	*t*	*df*	*p*	Cohen's *d*
das Holz [the wood]	**wird gehackt [is chopped]**	1.87	0.92	-14.40	38	< .001	2.31
das Sofa [the sofa]	**wird verladen [is loaded]**	1.90	0.85	-15.41	38	< .001	2.47
das Tier [the animal]	**wird geschossen [is shot]**	1.95	1.38	-9.31	38	< .001	1.49
das Fischfilet [the fish fillet]	**wird gegrillt [is grilled]**	2.00	1.12	-11.11	38	< .001	1.78
das Regal [the shelf]	**wird montiert [is mounted]**	2.00	1.15	-10.89	38	< .001	1.74
das Spielzeug [the toy]	**wird gesteuert [is steered]**	2.13	0.89	-13.08	38	< .001	2.09
das Motorboot [the motorboat]	**wird gesteuert [is steered]**	2.15	1.04	-11.09	38	< .001	1.78
das Brettspiel [the board game]	**wird eingebaut [is installed]**	2.18	1.07	-10.60	38	< .001	1.70
das Getränk [the drink]	**wird gezapft [is tapped]**	2.18	1.30	-8.78	38	< .001	1.41
das Paket [the parcel]	**wird angehoben [is lifted]**	2.36	1.14	-9.03	38	< .001	1.45
das Labyrinth [the labyrinth]	**wird vermessen [is measured]**	2.38	1.16	-8.69	38	< .001	1.39
das Handy [the cell phone]	**wird repariert [is repaired]**	2.46	1.37	-7.00	38	< .001	1.12
das Blatt [the sheet]	**wird zerstört [is destroyed]**	2.56	1.19	-7.55	38	< .001	1.21
das Mobiliar [the furniture]	**wird aufgebaut [is build up]**	2.62	1.35	-6.41	38	< .001	1.03
das Haus [the house]	**wird errichtet [is built]**	2.64	1.22	-6.93	38	< .001	1.11
das Fenster [the window]	**wird ausgemessen [is sized]**	2.67	1.24	-6.70	38	< .001	1.07
das Aquarium [the aquarium]	**wird aufgestellt [is positioned]**	2.77	1.27	-6.07	38	< .001	0.97
das Handtuch [the towel]	**wird beschmutzt [is dirtied]**	2.82	1.14	-6.44	38	< .001	1.03
das Glücksrad [the prize wheel]	**wird lackiert [is varnished]**	2.85	1.79	-4.04	38	< .001	0.65
das Holz [the wood]	**wird gelagert [is stored]**	2.87	1.24	-5.69	38	< .001	0.91
das Publikum [the audience]	**wird evakuiert [is evacuated]**	2.90	1.27	-5.41	38	< .001	0.87
das Kuscheltier [the cuddly toy]	**wird verbrannt [is burnt]**	2.92	1.20	-5.60	38	< .001	0.90
das Ladegerät [the charger]	**wird angeschlossen [is connected]**	3.00	1.12	-5.56	38	< .001	0.89
das Kleid [the dress]	**wird gestohlen [is stolen]**	3.00	1.12	-5.56	38	< .001	0.89

Note. 1 = stereotypically male, 7 = stereotypically female. The verb phrase's (marked in bold) and the anteceding noun phrase's translation to English are inserted in parentheses below.

(continued)

Table A26

Verb Phrases Considered Stereotypically Male According to a One-Sample t-Test Against the Scale Midpoint of 4 and the Noun Phrases That Anteceded the Verb Phrases in Pretest IV (continued)

Anteceding Noun Phrase	Verb Phrase	*M*	*SD*	*t*	*df*	*p*	Cohen's *d*
das Klavier [the piano]	**wird restauriert [is restored]**	3.00	1.36	-4.60	38	< .001	0.74
das Portemonnaie [the wallet]	**wird produziert [is produced]**	3.05	1.17	-5.07	38	< .001	0.81
das Radiergummi [the eraser]	**wird angefertigt [manufactured]**	3.13	1.06	-5.16	38	< .001	0.83
das Teleskop [the telescope]	**wird aufgestellt [is positioned]**	3.15	1.11	-4.75	38	< .001	0.76
das Regal [the shelf]	**wird justiert [is adjusted]**	3.21	1.51	-3.29	38	.002	0.53
das Raumschiff [the spaceship]	**wird erkundet [is explored]**	3.23	1.25	-3.86	38	< .001	0.62
das Eichhörnchen [the squirrel]	**wird verfolgt [is followed]**	3.26	1.31	-3.54	38	.001	0.57
das Wasser [the water]	**wird getestet [is tested]**	3.26	1.33	-3.49	38	.001	0.56
das Brot [the bread]	**wird gebrochen [is broken]**	3.28	1.08	-4.17	38	< .001	0.67
das Naturereignis [the natural event]	**wird prognostiziert [is forecasted]**	3.28	1.61	-2.79	38	.008	0.45
das Zimmer [the room]	**wird gestrichen [is painted]**	3.28	1.64	-2.74	38	.009	0.44
das Tablet [the tablet]	**wird upgedatet [is updated]**	3.31	1.28	-3.38	38	.002	0.54
das Teleskop [the telescope]	**wird umgestoßen [is knocked over]**	3.36	1.11	-3.60	38	.001	0.58
das Radio [the radio]	**wird reguliert [is regulated]**	3.38	1.50	-2.57	38	.014	0.41
das Pokerspiel [the poker game]	**wird verfolgt [is followed]**	3.44	1.05	-3.37	38	.002	0.54
das Museum [the museum]	**wird geräumt [is evacuated]**	3.44	1.27	-2.77	38	.009	0.44
das Sofa [the sofa]	**wird bepolstert [is cushioned]**	3.49	1.36	-2.36	38	.023	0.38
das Heft [the booklet]	**wird gedruckt [is printed]**	3.56	0.72	-3.79	38	.001	0.61
das Dokument [the document]	**wird gelöscht [is deleted]**	3.69	0.80	-2.40	38	.021	0.39

Note. 1 = stereotypically male, 7 = stereotypically female. The verb phrase's (marked in bold) and the anteceding noun phrase's translation to English are inserted in parentheses below.

Table A27

Verb Phrases Considered Stereotypically Female According to a One-Sample t-Test Against the Scale Midpoint of 4 and the Noun Phrases That Anteceded the Verb Phrases in Pretest IV

Anteceding Noun Phrase	**Verb Phrase**	*M*	*SD*	*t*	*df*	*p*	Cohen's *d*
das Heft [the booklet]	**wird ausgefüllt [is filled]**	4.36	0.90	2.48	38	.018	0.40
das Teleskop [the telescope]	**wird verwahrt [is kept safe]**	4.38	1.04	2.31	38	.027	0.37
das Museum [the museum]	**wird besucht [is visited]**	4.41	0.99	2.58	38	.014	0.41
das Raumschiff [the spaceship]	**wird gemieden [is avoided]**	4.44	0.85	3.20	38	.003	0.51
das Kind [the child]	**wird bespaßt [is entertained]**	4.44	1.21	2.25	38	.030	0.36
das Kuscheltier [the cuddly toy]	**wird gegeben [is given]**	4.46	0.85	3.38	38	.002	0.54
das Fachbuch [the reference book]	**wird gelesen [is read]**	4.54	0.91	3.68	38	.001	0.59
das Regal [the shelf]	**wird ausgewählt [is chosen]**	4.59	1.12	3.30	38	.002	0.53
das Xylophon [the xylophone]	**wird angehört [is listened]**	4.62	1.21	3.19	38	.003	0.51
das Handy [the cell phone]	**wird aufbewahrt [is stored]**	4.64	1.16	3.46	38	.001	0.55
das Publikum [the audience]	**wird wahrgenommen [is perceived]**	4.69	1.03	4.20	38	< .001	0.67
das Kind [the child]	**wird beschenkt [is made a present]**	4.72	0.97	4.61	38	< .001	0.74
das Brettspiel [the board game]	**wird organisiert [is organized]**	4.74	1.45	3.21	38	.003	0.51
das Getränk [the drink]	**wird umgerührt [is stirred up]**	4.77	0.96	5.01	38	< .001	0.80
das Sofa [the sofa]	**wird ausgesucht [is chosen]**	4.77	1.01	4.75	38	< .001	0.76
das Dokument [the document]	**wird vorgelesen [is read]**	4.77	1.18	4.07	38	< .001	0.65
das Glücksrad [the prize wheel]	**wird präsentiert [is presented]**	4.77	1.20	4.00	38	< .001	0.64
das Tablet [the tablet]	**wird gekauft [is bought]**	4.77	1.46	3.29	38	.002	0.53
das Klavier [the piano]	**wird ausgesucht [is chosen]**	4.82	0.91	5.61	38	< .001	0.90
das Fischfilet [the fish fillet]	**wird abgetaut [is defrosted]**	4.82	1.00	5.14	38	< .001	0.82
das Mobiliar [the furniture]	**wird ausgewählt [is chosen]**	4.85	0.99	5.35	38	< .001	0.86
das Radio [the radio]	**wird bewundert [is admired]**	4.97	1.22	4.97	38	< .001	0.80
das Paket [the parcel]	**wird bestellt [is ordered]**	4.97	1.31	4.65	38	< .001	0.75
das Spielzeug [the toy]	**wird desinfiziert [is disinfected]**	5.1	1.21	5.69	38	< .001	0.91
das Gift [the poison]	**wird versprüht [is sprayed]**	5.15	1.04	6.93	38	< .001	1.11
das Labyrinth [the labyrinth]	**wird gezeichnet [is drawn]**	5.18	1.19	6.19	38	< .001	0.99
das Wasser [the water]	**wird gefärbt [is colored]**	5.36	1.14	7.48	38	< .001	1.20

Note. 1 = stereotypically male, 7 = stereotypically female. The verb phrase's (marked in bold) and the anteceding noun phrase's translation to English are inserted in parentheses below.

(continued)

Table A27

Verb Phrases Considered Stereotypically Female According to a One-Sample t-Test Against the Scale Midpoint of 4 and the Noun Phrases That Anteceded the Verb Phrases in Pretest IV (continued)

Anteceding Noun Phrase	**Verb Phrase**	*M*	*SD*	*t*	*df*	*p*	Cohen's *d*
das Eichhörnchen [the squirrel]	**wird gewiegt [is cradled]**	5.36	1.39	6.12	38	< .001	0.98
das Brötchen [the roll]	**wird zubereitet [is prepared]**	5.44	1.07	8.37	38	< .001	1.34
das Ei [the egg]	**wird gekocht [is boiled]**	5.46	1.30	7.05	38	< .001	1.13
das Holz [the wood]	**wird angemalt [is painted]**	5.49	0.82	11.28	38	< .001	1.81
das Kleid [the dress]	**wird angezogen [is put on]**	5.51	1.19	7.94	38	< .001	1.27
das Pferd [the horse]	**wird geritten [is ridden]**	5.54	1.19	8.08	38	< .001	1.29
das Opossum [the opossum]	**wird angelächelt [is smiled at]**	5.56	1.07	9.12	38	< .001	1.46
das Tier [the animal]	**wird gestreichelt [is stroked]**	5.67	1.18	8.84	38	< .001	1.42
das Aquarium [the aquarium]	**wird gereinigt [is cleaned]**	5.67	1.28	8.11	38	< .001	1.30
das Album [the album]	**wird bemalt [is painted]**	5.69	0.95	11.12	38	< .001	1.78
das Haus [the house]	**wird aufgeräumt [is tidied]**	5.69	1.00	10.53	38	< .001	1.69
das Zimmer [the room]	**wird eingerichtet [is furnished]**	5.69	1.15	9.19	38	< .001	1.47
das Pferd [the horse]	**wird gestriegelt [is curried]**	5.69	1.22	8.68	38	< .001	1.39
das Brot [the bread]	**wird gebacken [is baked]**	5.72	1.10	9.76	38	< .001	1.56
das Kuscheltier [the cuddly toy]	**wird umarmt [is hugged]**	5.79	0.95	11.79	38	< .001	1.89
das Regal [the shelf]	**wird entstaubt [is dusted]**	5.79	1.03	10.88	38	< .001	1.74
das Plakat [the poster]	**wird verschönert [is beautified]**	6.08	0.90	14.41	38	< .001	2.34
das Plakat [the poster]	**wird ausgeschmückt [is embellished]**	6.13	0.80	16.60	38	< .001	2.66
das Sofa [the sofa]	**wird dekoriert [is decorated]**	6.13	0.89	14.87	38	< .001	2.38
das Sofa [the sofa]	**wird genäht [is sewn]**	6.21	1.01	13.71	38	< .001	2.20
das Handtuch [the towel]	**wird gebügelt [is ironed]**	6.21	1.06	13.04	38	< .001	2.09
das Fenster [the window]	**wird geputzt [is cleaned]**	6.23	0.84	16.55	38	< .001	2.65
das Regal [the shelf]	**wird dekoriert [is decorated]**	6.23	0.99	14.13	38	< .001	2.26
das Kind [the child]	**wird bemuttert [mothered]**	6.74	0.68	25.29	38	< .001	4.05

Note. 1 = stereotypically male, 7 = stereotypically female. The verb phrase's (marked in bold) and the anteceding noun phrase's translation to English are inserted in parentheses below.

Table A28

Verb Phrases Considered Stereotypically Gender-Neutral According to a One-Sample t-Test Against the Scale Midpoint of 4 and the Noun Phrases That Anteceded the Verb Phrases in Pre-test IV

Anteceding Noun Phrase	**Verb Phrase**	*M*	*SD*	*t*	*df*	*p*	Cohen's *d*
das Album [the album]	**wird erkundet [is explored]**	3.54	1.47	-1.97	38	.057	0.32
das Ei [the egg]	**wird aufgeschlagen [is cracked]**	3.59	1.33	-1.92	38	.062	0.31
das Gift [the poison]	**wird erforscht [is researched]**	3.69	1.44	-1.34	38	.189	0.21
das Raumschiff [the spaceship]	**wird gesichtet [is sighted]**	3.82	0.97	-1.16	38	.255	0.19
das Opossum [the opossum]	**wird verspottet [is mocked]**	3.85	1.37	-0.70	38	.487	0.11
das Motorboot [the motorboat]	**wird betreten [is entered]**	3.87	0.73	-1.09	38	.281	0.18
das Brötchen [the roll]	**wird aufgeschnitten [is sliced open]**	3.90	0.97	-0.66	38	.512	0.11
das Pokerspiel [the poker game]	**wird erklärt [is explained]**	3.90	1.33	-0.48	38	.634	0.08
das Motorboot [the motorboat]	**wird zerkrazt [is scratched]**	3.92	1.27	-0.38	38	.706	0.06
das Ladegerät [the charger]	**wird aufgerollt [is rolled]**	3.97	1.56	-0.10	38	.919	0.02
das Naturereignis [the natural event]	**wird gefürchtet [is feared]**	4.13	1.79	0.45	38	.658	0.07
das Fachbuch [the reference book]	**wird eingesteckt [is packed]**	4.18	1.14	0.98	38	.333	0.16
das Pokerspiel [the poker game]	**wird verachtet [is despised]**	4.21	1.34	0.96	38	.346	0.15
das Portemonnaie [the wallet]	**wird gefunden [is found]**	4.23	1.27	1.14	38	.262	0.18
das Blatt [the sheet]	**wird entworfen [is designed]**	4.31	1.54	1.25	38	.220	0.20
das Pferd [the horse]	**wird verarztet [is doctored]**	4.31	1.56	1.23	38	.225	0.20
das Xylophon [the xylophone]	**wird gestimmt [is tuned]**	4.33	1.06	1.97	38	.057	0.32
das Kleid [the dress]	**wird komplimentiert [is complimented]**	4.41	1.31	1.95	38	.058	0.31
das Radiergummi [the eraser]	**wird eingepackt [is packed]**	4.44	1.67	1.63	38	.111	0.26

Note. 1 = stereotypically male, 7 = stereotypically female. The verb phrase's (marked in bold) and the anteceding noun phrase's translation to English are inserted in parentheses below.

A Experiment 2 (Main Verbs - Humans) – Verbal Stimuli: The Experimental Sentences

Table A29

The Experimental Sentences of Experiment 2 in Tuples, Their Classification in Terms of Speaker Voice, Main Verb Gender-Stereotypicality, and NP2 Gender, and Their Exact Word On- and Offsets

Tuple	Speaker Voice	Main Verb Gender	NP2 Gender	Experimental Sentence	NP1 Off	V1 On	V2 On	V2 Off	Prep. On	Prep. Off	NP2 On	NP2 Off
1	Male	Male	Male	Das Fenster wird **ausgemessen** von dem Bauarbeiter.	908	1525	1757	2843	3362	3750	4001	5048
	Female			[The window is **sized** by the construction worker.]	1119	1547	1978	3030	3899	4193	4194	5354
	Male	Female		Das Fenster wird **geputzt** von dem Bauarbeiter.	975	1526	1758	2492	3363	3724	4001	5129
	Female			[The window is **cleaned** by the construction worker.]	1110	1550	1975	2677	3896	4302	4303	5529
2	Male	Male	Male	Das Tablet wird **upgedate**t von dem Türsteher.	889	1308	1693	2674	2991	3370	3766	4841
	Female			[The tablet is **updated** by the bouncer.]	919	1355	1797	3002	3216	3633	3838	5054
	Male	Female		Das Tablet wird **gekauft** von dem Türsteher.	860	1311	1692	2477	2992	3468	3764	4790
	Female			[The tablet is **bought** by the bouncer.]	829	1352	1758	2559	3201	3637	3836	4930
3	Male	Male	Female	Das Brot wird **gebrochen** von der Flugbegleiterin.	770	1481	1742	2426	2903	3396	3785	5044
	Female			[The bread is **broken** by the stewardess.]	950	1380	1914	2787	3396	3899	3973	5522
	Male	Female		Das Brot wird **gebacken** von der Flugbegleiterin.	746	1483	1745	2379	2907	3358	3781	4975
	Female			[The bread is **baked** by the stewardess.]	977	1379	1912	2711	3394	3759	3968	5512
4	Male	Male	Female	Das Heft wird **gedruckt** von der Parfümeriefachverkäuferin.	740	1182	1464	2089	2588	3066	3403	5368
	Female			[The booklet is **printed** by the perfumery shop assistant.]	821	1325	1879	2571	3097	3584	3724	6254
	Male	Female		Das Heft wird **dekoriert** von der Parfümeriefachverkäuferin.	734	1180	1472	2229	2586	3050	3405	5237
	Female			[The booklet is **decorated** by the perfumery shop assistant.]	852	1374	1879	2758	3095	3587	3722	6131
5	Male	Male	Female	Das Getränk wird **gezapft** von der Flugbegleiterin.	901	1581	1840	2593	2925	3400	3749	4950
	Female			[The drink is **tapped** by the stewardess.]	1053	1460	1900	2701	3223	3617	3739	5226
	Male	Female		Das Getränk wird **umgerührt** von der Flugbegleiterin.	930	1582	1915	2621	2923	3325	3722	4965
	Female			[The drink is **stirred up** by the stewardess.]	1149	1460	1901	2831	3223	3636	3738	5226

Note. On/Off = On- and offsets = the time in ms from sentence onset, the onset of NP1 was 0ms for all sentences and is thus not listed here for clarity of presentation. NP = Noun Phrase, V = Verb (Main Verb = V2, marked in bold), Prep. = Preposition. The sentence's translation to English is inserted in parentheses below.

(continued)

Table A29

The Experimental Sentences of Experiment 2 in Tuples, Their Classification in Terms of Speaker Voice, Main Verb Gender-Stereotypicality, and NP2 Gender, and Their Exact Word On- and Offsets (continued)

Tuple	Speaker Voice	Main Verb Gender	NP2 Gender	Experimental Sentence	NP1 Off	V1 On	V2 On	V2 Off	Prep. On	Prep. Off	NP2 On	NP2 Off
6	Male	Male	Female	Das Regal wird **justiert** von der Parfümeriefachverkäuferin.	898	1474	1755	2456	2979	3450	3682	5548
	Female			[The shelf is **adjusted** by the perfumery shop assistant.]	1042	1484	1833	2740	3399	3448	4018	6615
	Male	Female		Das Regal wird **entstaubt** von der Parfümeriefachverkäuferin.	898	1477	1809	2673	2979	3329	3683	5530
	Female			[The shelf is **dusted off** by the perfumery shop assistant.]	1000	1484	1895	2877	3395	3851	4053	6615
7	Male	Male	Female	Das Xylophon wird **angehört** von der Hausfrau.	1227	1713	1955	2686	3046	3570	3891	4807
	Female			[The xylophone is **listened** by the housewife.]	1167	1512	2039	2845	3503	3911	4061	5029
	Male	Female		Das Xylophon wird **desinfiziert** von der Hausfrau.	1286	1713	1955	3027	3046	3504	3891	4849
	Female			[The xylophone is **disinfected** by the housewife.]	1140	1512	2041	3262	3503	3864	4061	5057
8	Male	Male	Male	Das Blatt wird **zerstört** von dem Feuerwehrmann.	789	1374	1645	2407	2982	3475	3693	4825
	Female			[The sheet is **destroyed** by the firefighter.]	907	1577	2175	3142	3918	4289	4440	5578
	Male	Female		Das Blatt wird **verschönert** von dem Feuerwehrmann.	834	1374	1646	2470	2984	3392	3695	4872
	Female			[The sheet is **beautified** by the firefighter.]	856	1578	2176	3146	3917	4330	4440	5578
9	Male	Male	Female	Das Radiergummi wird **angefertigt** von der Kosmetikerin.	1139	1598	1867	2877	3005	3395	3645	4894
	Female			[The eraser is **manufactured** by the cosmetician.]	1159	1526	1897	2962	3363	3761	3936	5254
	Male	Female		Das Radiergummi wird **wahrgenommen** von der Kosmetikerin.	1181	1595	1868	2710	3004	3460	3647	4878
	Female			[The eraser is **perceived** by the cosmetician.]	1123	1524	1903	2904	3364	3788	3936	5318
10	Male	Male	Male	Das Paket wird **angehoben** von dem Soldaten.	963	1382	1657	2443	2891	3262	3328	4316
	Female			[The parcel is **lifted** by the soldier.]	1002	1784	2137	3114	3393	3842	3976	5138
	Male	Female		Das Paket wird **bestellt** von dem Soldaten.	902	1382	1658	2363	2891	3262	3328	4316
	Female			[The parcel is **ordered** by the soldier.]	1045	1784	2138	2899	3392	3857	3977	5232
11	Male	Male	Female	Das Plakat wird **aufgebaut** von der Friseurin.	1009	1512	1700	2562	2669	3136	3457	4508
	Female			[The poster is **built up** by the hair dresser.]	1143	1751	2223	3259	3768	4267	4395	5574
	Male	Female		Das Plakat wird **ausgeschmückt** von der Friseurin.	992	1512	1702	2620	2669	3136	3457	4508
	Female			[The poster is **embellished** by the hair dresser.]	1143	1751	2223	3363	3768	4187	4395	5509

Note. On/Off = On- and offsets = the time in ms from sentence onset, the onset of NP1 was 0ms for all sentences and is thus not listed here for clarity of presentation. NP = Noun Phrase, V = Verb (Main Verb = V2, marked in bold), Prep. = Preposition. The sentence's translation to English is inserted in parentheses below.

(continued)

Table A29

The Experimental Sentences of Experiment 2 in Tuples, Their Classification in Terms of Speaker Voice, Main Verb Gender-Stereotypicality, and NP2 Gender, and Their Exact Word On- and Offsets (continued)

Tuple	Speaker Voice	Main Verb Gender	NP2 Gender	Experimental Sentence	NP1 Off	V1 On	V2 On	V2 Off	Prep. On	Prep. Off	NP2 On	NP2 Off
12	Male	Male	Male	Das Ei wird gestohlen von dem Mechatroniker.	815	1302	1542	2348	2665	3089	3445	4646
	Female			[The egg is **stolen** by the mechatronics engineer.]	855	1178	1555	2453	2819	3204	3308	4551
	Male	Female		Das Ei wird gekocht von dem Mechatroniker.	814	1302	1541	2226	2663	3070	3445	4693
	Female			[The egg is **boiled** by the mechatronics engineer.]	801	1177	1555	2357	2819	3186	3308	4551
13	Male	Male	Female	Das Portemonnaie wird **produziert** von der Hausfrau.	1055	1574	1861	2643	3146	3569	3819	4793
	Female			[The wallet is **produced** by the housewife.]	1197	1680	2161	3026	3470	3875	4150	5185
	Male	Female		Das Portemonnaie wird **aufbewahrt** von der Hausfrau.	1154	1577	1855	2765	3146	3596	3815	4731
	Female			[The wallet is **stored** by the housewife.]	1174	1680	2161	3076	3470	3900	4149	5186
14	Male	Male	Male	Das Tier wird **geschossen** von dem Türsteher.	863	1502	1775	2450	3099	3463	3764	4770
	Female			[The animal is **shot** by the bouncer.]	852	1295	1806	2708	3150	3589	3973	5119
	Male	Female		Das Tier wird **gestreichelt** von dem Türsteher.	881	1500	1775	2621	3094	3582	3769	4814
	Female			[The animal is **stroked** by the bouncer.]	869	1295	1807	2732	3149	3548	3975	5120
15	Male	Male	Male	Das Dokument wird **gelöscht** von dem Goldgräber.	1058	1500	1778	2249	2780	3195	3413	4513
	Female			[The document is **deleted** by the gold digger.]	1145	1807	2363	3054	3652	4082	4179	5270
	Male	Female		Das Dokument wird **vorgelesen** von dem Goldgräber.	1043	1500	1771	2717	2784	3153	3414	4504
	Female			[The document is **read** by the gold digger.]	1155	1807	2363	3338	3695	4050	4180	5268
16	Male	Male	Male	Das Radio wird **reguliert** von dem Metzger.	942	1309	1705	2408	2788	3223	3533	4399
	Female			[The radio is **regulated** by the butcher.]	1049	1422	1849	2703	3193	3591	3833	4844
	Male	Female		Das Radio wird **bewundert** von dem Metzger.	981	1313	1700	2409	2760	3168	3491	4345
	Female			[The radio is **admired** by the butcher.]	1044	1424	1848	2709	3194	3540	3833	4845
17	Male	Male	Female	Das Aquarium wird **aufgestellt** von der Floristin.	1282	1835	2105	3049	3309	3778	4023	5009
	Female			[The aquarium is **positioned** by the florist.]	1190	1595	2151	3143	3707	4119	4266	5432
	Male	Female		Das Aquarium wird **gereinigt** von der Floristin.	1282	1835	2139	3076	3300	3755	3943	5001
	Female			[The aquarium is **cleaned** by the florist.]	1150	1589	2150	3102	3711	4110	4262	5432

Note. On/Off = On- and offsets = the time in ms from sentence onset, the onset of NP1 was 0ms for all sentences and is thus not listed here for clarity of presentation. NP = Noun Phrase, V = Verb (Main Verb = V2, marked in bold), Prep. = Preposition. The sentence's translation to English is inserted in parentheses below.

(continued)

Table A29

The Experimental Sentences of Experiment 2 in Tuples, Their Classification in Terms of Speaker Voice, Main Verb Gender-Stereotypicality, and NP2 Gender, and Their Exact Word On- and Offsets (continued)

Tuple	Speaker Voice	Main Verb Gender	NP2 Gender	Experimental Sentence	NP1 Off	V1 On	V2 On	V2 Off	Prep. On	Prep. Off	NP2 On	NP2 Off
18	Male	Male	Male	Das Publikum wird **evakuiert** von dem Bauarbeiter.	1001	1400	1665	2417	2691	3116	3705	4812
	Female			[The audience is **evacuated** by the construction worker.]	1116	1452	1920	2829	3289	3718	3835	5096
	Male	Female		Das Publikum wird **bemuttert** von dem Bauarbeiter.	1009	1400	1664	2305	2691	3139	3704	4801
	Female			[The audience is **cared** by the construction worker.]	1121	1456	1922	2780	3289	3689	3835	5101
19	Male	Male	Female	Das Zimmer wird **restauriert** von der Zahnarzthelferin.	867	1245	1475	2264	2681	3110	3301	4761
	Female			[The room is **restored** by the dentist assistant.]	838	1051	1451	2427	2797	3214	3325	4852
	Male	Female		Das Zimmer wird **eingerichtet** von der Zahnarzthelferin.	867	1245	1475	2279	2671	3106	3304	4732
	Female			[The room is **furnished** by the dentist assistant.]	834	1052	1447	2439	2776	3167	3324	4852
20	Male	Male	Male	Das Museum wird **geräumt** von dem Feuerwehrmann.	973	1544	1803	2588	3015	3562	4003	5014
	Female			[The museum is **evacuated** by the firefighter.]	994	1276	1739	2478	3058	3441	3676	4817
	Male	Female		Das Museum wird **beschenkt** von dem Feuerwehrmann.	981	1541	1852	2533	3015	3562	4003	5015
	Female			[The museum is **made a present** by the firefighter.]	1027	1275	1739	2512	3055	3489	3675	4812
21	Male	Male	Female	Das Handtuch wird **beschmutzt** von der Kosmetikerin.	915	1410	1701	2424	3044	3522	3770	5053
	Female			[The towel is **dirtied** by the cosmetician.]	896	1316	1838	2569	3344	3686	3856	5199
	Male	Female		Das Handtuch wird **gebügelt** von der Kosmetikerin.	915	1410	1722	2438	3003	3500	3770	5025
	Female			[The towel is **ironed** by the cosmetician.]	858	1317	1838	2605	3327	3735	3855	5200
22	Male	Male	Male	Das Gift wird **erforscht** von dem Soldaten.	828	1338	1589	2266	2653	3155	3438	4442
	Female			[The poison **researched** by the soldier.]	826	1383	1813	2563	3044	3506	3653	4733
	Male	Female		Das Gift wird **versprüht** von dem Soldaten.	828	1338	1589	2258	2639	3155	3422	4447
	Female			[The poison **sprayed** by the soldier.]	797	1383	1815	2588	3039	3470	3654	4732
23	Male	Male	Female	Das Album wird **verbrannt** von der Friseurin.	980	1706	1954	2632	3116	3525	3814	4855
	Female			[The album is **burnt** by the hair dresser.]	914	1191	1653	2475	2992	3446	3553	4589
	Male	Female		Das Album wird **bemalt** von der Friseurin.	980	1700	1976	2628	3139	3643	3814	4870
	Female			[The album is **painted** by the hair dresser.]	931	1191	1652	2426	2992	3430	3553	4588

Note. On/Off = On- and offsets = the time in ms from sentence onset, the onset of NP1 was 0ms for all sentences and is thus not listed here for clarity of presentation. NP = Noun Phrase, V = Verb (Main Verb = V2, marked in bold), Prep. = Preposition. The sentence's translation to English is inserted in parentheses below.

(continued)

Table A29

The Experimental Sentences of Experiment 2 in Tuples, Their Classification in Terms of Speaker Voice, Main Verb Gender-Stereotypicality, and NP2 Gender, and Their Exact Word On- and Offsets (continued)

Tuple	Speaker Voice	Main Verb Gender	NP2 Gender	Experimental Sentence	NP1 Off	V1 On	V2 On	V2 Off	Prep. On	Prep. Off	NP2 On	NP2 Off
24	Male	Male	Male	Das Fischfilet wird **gegrillt** von dem Mechatroniker.	1110	1788	2084	2669	3352	3752	4142	5342
	Female			[The fish fillet is **roasted** by the mechatronics engineer.]	1182	1411	1879	2666	3160	3550	3664	4941
	Male	Female		Das Fischfilet wird **abgetaut** von dem Mechatroniker	1118	1791	1998	2749	3352	3761	4144	5345
	Female			[The fish fillet is **defrosted** by the mechatronics engineer.]	1134	1410	1880	2705	3161	3543	3665	4946
25	Male	Male	Male	Das Labyrinth wird **vermessen** von dem Türsteher.	1022	1613	1836	2600	3148	3710	1469	5153
	Female			[The labyrinth is **measured** by the bouncer.]	1073	1458	1984	2804	3138	3563	3686	4717
	Male	Female		Das Labyrinth wird **gezeichnet** von dem Türsteher.	1022	1614	1857	2663	3145	3630	4165	5151
	Female			[The labyrinth is **drawn** by the bouncer.]	1109	1462	1985	2840	3140	3555	3689	4716
26	Male	Male	Female	Das Naturereignis wird **gesteuert** von der Flugbegleiterin.	1335	1736	1975	2692	3296	3740	4042	5404
	Female			[The natural event is **controlled** by the stewardess.]	1478	1829	2301	3038	3585	3948	4075	5473
	Male	Female		Das Naturereignis wird **präsentiert** von der Flugbegleiterin.	1337	1736	1952	2756	3310	3727	4049	5407
	Female			[The natural event is **presented** by the stewardess.]	1487	1834	2303	3294	3545	3908	4077	5480
27	Male	Male	Male	Das Haus wird **errichtet** von dem Goldgräber.	668	1000	1172	1903	2430	2891	3263	4317
	Female			[The house is **raised** by the gold digger.]	820	1074	1566	2335	2888	3295	3363	4423
	Male	Female		Das Haus wird **aufgeräumt** von dem Goldgräber.	668	1000	1172	2034	2430	2891	3263	4317
	Female			[The house is **tidied** by the gold digger.]	838	1074	1572	2505	2888	3311	3363	4411
28	Male	Male	Male	Das Opossum wird **verspottet** von dem Metzger.	1000	1440	1621	2421	3006	3407	3782	4509
	Female			[The opossum is **mocked** by the butcher.]	1059	1467	1811	2781	3184	3510	3624	4425
	Male	Female		Das Opossum wird **angelächelt** von dem Metzger.	1000	1435	1680	2506	3006	3407	3797	4509
	Female			[The opossum is **smiled at** by the butcher.]	1066	1467	1823	2714	3171	3536	3624	4517
29	Male	Male	Female	Das Wasser wird **getestet** von der Floristin.	845	1464	1702	2494	2855	3306	3785	4753
	Female			[The water is **tested** by the florist.]	858	1147	1685	2530	2785	3191	3258	4252
	Male	Female		Das Wasser wird **gefärbt** von der Floristin.	839	1455	1715	2508	2868	3339	3799	4761
	Female			Das Wasser wird **dyed** von der Floristin	885	1147	1685	2473	2784	3175	3258	4276

Note. On/Off = On- and offsets = the time in ms from sentence onset, the onset of NP1 was 0ms for all sentences and is thus not listed here for clarity of presentation. NP = Noun Phrase, V = Verb (Main Verb = V2, marked in bold), Prep. = Preposition. The sentence's translation to English is inserted in parentheses below.

(continued)

Table A29

The Experimental Sentences of Experiment 2 in Tuples, Their Classification in Terms of Speaker Voice, Main Verb Gender-Stereotypicality, and NP2 Gender, and Their Exact Word On- and Offsets (continued)

Tuple	Speaker Voice	Main Verb Gender	NP2 Gender	Experimental Sentence	NP1 Off	V1 On	V2 On	V2 Off	Prep. On	Prep. Off	NP2 On	NP2 Off
30	Male	Male	Female	Das Eichhörnchen wird **verfolgt** von der Zahnarzthelferin.	1107	1754	1919	2733	3262	3771	4130	5529
	Female			[The squirrel is **followed** by the dental assistant.]	1361	1716	2222	3071	3293	3741	3921	5578
	Male	Female		Das Eichhörnchen wird **gewiegt** von der Zahnarzthelferin.	1107	1754	1997	2720	3262	3771	4159	5529
	Female			[The squirrel is **cradled** by the dental assistant.]	1313	1726	2222	2832	3293	3702	3886	5489
31	Male	Male	Male	Das Sofa wird **verladen** von dem Bauarbeiter.	839	1472	1649	2477	2854	3180	3593	4679
	Female			[The sofa is **loaded** by the construction worker.]	1087	1418	2013	2860	3056	3448	3622	4903
	Male	Female		Das Sofa wird **genäht** von dem Bauarbeiter.	839	1472	1649	2315	2854	3180	3589	4681
	Female			[The sofa is **sewn** by the construction worker.]	1017	1418	2013	2652	3056	3459	3622	4892
32	Male	Male	Female	Das Brettspiel wird **eingebaut** von der Hausfrau.	1052	1463	1690	2536	3047	3544	3873	4779
	Female			[The board game is **installed** by the housewife.]	1226	1633	2035	2947	3411	3755	3902	4831
	Male	Female		Das Brettspiel wird **ausgesucht** von der Hausfrau.	1052	1463	1662	2579	3047	3544	3873	4779
	Female			[The board game is **chosen** by the housewife.]	1309	1633	2035	3041	3411	3816	3913	4860

Note. On/Off = On- and offsets = the time in ms from sentence onset, the onset of NP1 was 0ms for all sentences and is thus not listed here for clarity of presentation. NP = Noun Phrase, V = Verb (Main Verb = V2, marked in bold), Prep. = Preposition. The sentence's translation to English is inserted in parentheses below.

A Experiment 2 (Main Verbs - Humans) – Verbal Stimuli: The Filler Sentences

Table A30

The Filler Sentences of Experiment 2, Their Classifications in Terms of NP1 and NP2 Gender-Stereotypicality, the NP2's Grammatical Gender, and the Corresponding Questions Related to Either the Content of the Filler Sentences or to the Pictures

Item	NP1 Gender	NP2 Gender	Filler Sentence	Question following the filler sentence	Correct Answer	Question's Focus
F1	Male	Male	Das Auto wird gemietet von dem Mechatroniker. [The car is rent by the mechatronics engineer (male form).]	Trug der Mechatroniker blaue Kleidung? [Did the mechatronics engineer (male form) wear blue clothes?]	Yes	Picture
F2	Male	Female	Das Steak wird gebraten von der Hausfrau. [The steak is roasted by the housewife (female form).]	Wird das Steak gebraten von der Hausfrau? [Is the steak roasted by the housewife (female form)?]	Yes	Sentence
F3	Male	Male	Das Zelt wird bewohnt von der Mechatroniker**in**. [The tent is inhabited by the mechatronics engineer **(female form)**.]	War das Zelt gelb? [Was the tent yellow?]	No	Picture
F4	Male	Female	Das Lenkrad wird gedreht von dem Haus**mann**. [The steering wheel is turned by the househusband **(male form)**.]	Wird das Lenkrad gedreht von dem Haus**mann**? [Is the steering wheel turned by the house**husband (male form)**?]	Yes	Sentence
F5	Female	Male	Das Haaröl wird weggeschmissen von dem Türsteher. [The hair oil is thrown away by the bouncer (male form).]	Trug der Türsteher einen Ledermantel? [Did the bouncer (male form) wear a black leather coat?]	No	Picture
F6	Female	Female	Das Pferd wird gefilmt von der Flugbegleiterin. [The horse is filmed by the stewardess (female form).]	Wird das Pferd gefüttert von der Flugbegleiterin? [Is the horse fed by the stewardess (female form)?]	No	Sentence
F7	Female	Male	Das Haarband wird gewechselt von der Türsteher**in**. [The hair band is changed by the bouncer **(female form).**]	War das Haarband grün? [Was the hair band green?]	No	Picture
F8	Female	Female	Das Spülmittel wird verstaut von dem Flugbegleit**er**. [The detergent is stored by the stewar**d (male form)**.]	Wird das Spülmittel verstaut von dem Flugbegleit**er**? (Is the detergent stored by the stewar**d (male form)?]**	Yes	Sentence
F9	Male	Male	Das Fass wird gefüllt von dem Goldgräber. [The cask is filled by the gold digger **(male form)**?	War das Fass aus Holz? [Was the cask wooden?]	Yes	Picture
F10	Male	Female	Das Bier wird verschüttet von der Parfümeriefachverkäuferin. [The beer is spilled by the perfumery shop assistant (female form).]	Wird das Bier verschüttet von dem Fris**eur**? [Is the beer spilled by the hair dresser **(male form)**?]	No	Sentence
F11	Male	Male	Das Motorrad wird abgeholt von der Goldgräber**in**. [The motorbike is picked up by the gold digger **(female form)**.]	Wird das Motorrad gefahren von der Goldgräber**in**? [Is the motorbike driven by the gold digger **(female form)**?]	No	Sentence
F12	Male	Female	Das Eisen wird hingelegt von dem Parfümeriefachverkäuf**er**. [Is the iron laid down by the perfumery shop assistant **(male form)**?]	Trug die Parfümeriefachverkäuferin ein oranges Top? [Did the perfumery shop assistant (female form) wear an orange top?]	Yes	Picture

Note. NP1 and NP2 Gender = an NP1 object's and a NP2 profession's gender-stereotypicality. Deviations between a NP2 profession's gender-stereotypicality and its grammatical gender are marked in bold. Verb phrases used in the filler sentences were stereotypically gender-neutral. The sentence's and the question's translation to English are inserted in parentheses below.

(continued)

Table A30

The Filler Sentences of Experiment 2, Their Classifications in Terms of NP1 and NP2 Gender-Stereotypicality, the NP2's Grammatical Gender, and the Corresponding Questions Related to Either the Content of the Filler Sentences or to the Pictures (continued)

Item	NP1 Gender	NP2 Gender	Filler Sentence	Question following the filler sentence	Correct Answer	Question's Focus
F13	Female	Male	Das Ballett wird angesehen von dem Metzger. [The ballet is watched by the butcher (male form).]	Wird das Ballett angesehen von dem Handwerker? [Is the ballet watched by the craftsman (male form)?]	No	Sentence
F14	Female	Female	Das Schmuckstück wird bestaunt von der Floristin. [The jewelry is gazed by the florist (female form).]	War das abgebildete Schmuckstück eine Kette? [Was the depicted jewelry a necklace?]	No	Picture
F15	Female	Male	Das Kleid wird entworfen von der Metzger**in**. [The dress is designed by the butcher **(female form)**.]	War das Kleid blau? [Was the dress blue?]	Yes	Picture
F16	Female	Female	Das Porzellan wird zerbrochen von dem Floristen. [The porcelain is broken by the florist **(male form)**.]	Wird das Porzellan zerbrochen von dem Florist**en**? [Is the porcelain broken by the florist **(male form)**?]	Yes	Sentence
F17	Male	Male	Das Werkzeug wird gesucht von dem Bauarbeiter. [The tool is searched by the construction worker (male form).]	War das abgebildete Werkzeug ein Hammer? [Was the depicted tool a hammer?]	No	Picture
F18	Male	Female	Das Motorboot wird abgestellt von der Zahnarzthelferin. [The motorboat is turned off by the dental assistant (female form).]	Wird das Motorboot abgestellt von dem Zahnarzt? [Is the motorboat turned off by the dentist (male form)?]	No	Sentence
F19	Male	Male	Das Metermaß wird zugeschnitten von der Bauarbeiter**in**. [The tape measure is cut by the construction worker **(female form)**.]	Trug die Bauarbeiter**in** einen Helm? [Did the construction worker **(female form)** wear a helmet?]	Yes	Picture
F20	Male	Female	Das Raumschiff wird studiert von dem Zahnarzthelf**er**. [The spaceship is studied by the dental assistant **(male form)**.]	Wird das Raumschiff studiert von dem Zahnarzthelf**er**? [Is the spaceship studied by the dental assistant **(male form)**?]	Yes	Sentence
F21	Female	Male	Das Armband wird verschickt von dem Feuerwehrmann. [The bracelet is sent by the firefighter (male form).]	Wird das Holzspiel verschickt von dem Feuerwehrmann? [Is the wooden toy sent by the firefighter (male form)?]	No	Sentence
F22	Female	Female	Das Spülbecken wird benutzt von der Kosmetikerin. [The sink is used by the cosmetician (female form).]	Hatte die Kosmetikerin ein blaues Kleid an? [Did the cosmetician (female form) wear a blue dress?]	Yes	Picture
F23	Female	Male	Das Baby wird gesehen von der Feuerwehr**frau.** [The baby is seen by the firefighter **(female form)**.]	Wird das Baby gefüttert von der Feuerwehr**frau**? [Is the baby fed by the firefighter **(female form)**?]	No	Sentence
F24	Female	Female	Das Ballkleid wird aufgehängt von dem Kosmetik**er**. [The ball dress is hung up by the cosmetician **(male form)**.]	War das Ballkleid weiß? [Was the ball dress white?]	Yes	Picture
F25	Male	Male	Das Kabel wird übersehen von dem Soldaten. [The cable is overlooked by the soldier (male form).]	Wird das Kabel übersehen von dem Bäcker? [Is the cable overlooked by the baker (male form)?]	No	Sentence
F26	Male	Female	Das Rasiermesser wird entsorgt von der Friseurin. [The razor is disposed by the hair dresser (female form).]	War das Rasiermesser aufgeklappt? [Was the razor flipped open?]	Yes	Picture

Note. NP1 and NP2 Gender = an NP1 object's and a NP2 profession's gender-stereotypicality. Deviations between a NP2 profession's gender-stereotypicality and its grammatical gender are marked in bold. Verb phrases used in the filler sentences were stereotypically gender-neutral. The sentence's and the question's translation to English are inserted in parentheses below.

(continued)

Table A30

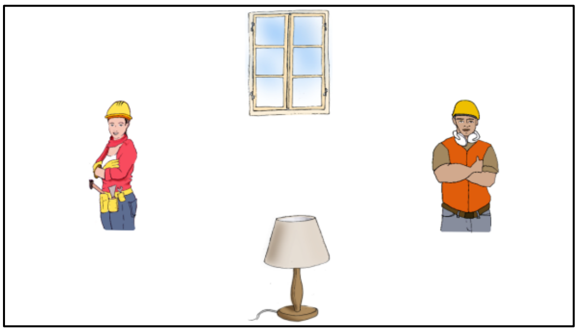
The Filler Sentences of Experiment 2, Their Classifications in Terms of NP1 and NP2 Gender-Stereotypicality, the NP2's Grammatical Gender, and the Corresponding Questions Related to Either the Content of the Filler Sentences or to the Pictures (continued)

Item	NP1 Gender	NP2 Gender	Filler Sentence	Question following the filler sentence	Correct Answer	Question's Focus
F27	Male	Male	Das Unternehmen wird eröffnet von der Soldat**in**. [The company is opened by the soldier **(female form)**.]	Wird das Unternehmen eröffnet von der Soldat**in**? [Is the company opened by the soldier **(female form)**?]	Yes	Sentence
F28	Male	Female	Das Hemd wird besorgt von dem Fris**eur**. [The shirt is gotten by the hair dresser **(male form)**.]	War das Hemd rot? [Was the shirt red?]	No	Picture
F29	Female	Male	Das Kochfeld wird eingeschaltet von dem Mechatroniker. [The stove is turned on by the mechatronics engineer (male form).]	Wird das Kochfeld ausgebaut von dem Mechatroniker? [Is the stove removed by the mechatronics engineer (male form)?]	No	Sentence
F32	Female	Female	Das Rezept wird beschrieben von dem Haus**mann**. [The receipt is described by the house**husband (male form)**.]	Wird das Rezept beschrieben von der Hausfrau? [Is the receipt descriped by the housewife (female form)?]	No	Sentence
F33	Male	Male	Das Fußballfeld wird gesichtet von dem Türsteher. [The soccer field is sighted by the bouncer (male form).]	War das Fußballfeld blau? [Was the soccer field blue?]	No	Picture
F34	Male	Female	Das Stadion wird besichtigt von der Flugbegleiterin. [The stadium is visited by the stewardess (female form).]	War in dem Stadion die amerikanische Flagge abgebildet? [Was the American flag depicted in the stadium?]	Yes	Picture
F35	Male	Male	Das Bauwerk wird betrachtet von der Türsteher**in**. [The building is watched by the bouncer **(female form)**.]	Wird das Haus betrachtet von der Türsteher**in**? [Is the house watched by the bouncer **(female form)**?]	No	Sentence
F36	Male	Female	Das Teleskop wird verwahrt von dem Flugbegleiter. [The telescope is kept safe by the stewar**d (male form)**.]	Hatte das Teleskop drei Standbeine? [Did the telescope have three pillars?]	Yes	Picture
F37	Female	Male	Das Bett wird geschoben von dem Goldgräber. [The bed is pushed by the gold digger (male form).]	War die Bettwäsche in rot und weiß kariert? [Was the linen red and white checkered?]	Yes	Picture
F38	Female	Female	Das Puppenhaus wird bestückt von der Parfümeriefachverkäuferin. [The doll house is equipped by the perfumery shop assistant (female from).]	Wird das Puppenhaus zerstört von der Parfümeriefachverkäuferin? [Is the doll house destroyed by the perfumery shop assistant (female form)?]	No	Sentence
F39	Female	Male	Das Sekretariat wird gemieden von der Goldgräber**in**. [The secretariat is avoided by the gold digger **(female form)**.]	War auf dem Bild des Sekretariats ein Computerbildschirm zu sehen? [Was there a pc screen in the depicted secretariat?]	Yes	Picture
F40	Female	Female	Das Make-Up wird genutzt von dem Parfümeriefachverkäuf**er**. [The make-up is used by the perfumery shop assistant **(male form)**.]	Wird das Make-Up weggeschmissen von dem Parfümeriefachverkäuf**er**? [Is the make-up thrown away by the perfumery shop assistant **(male form)**?]	No	Sentence

Note. NP1 and NP2 Gender = an NP1 object's and a NP2 profession's gender-stereotypicality. Deviations between a NP2 profession's gender-stereotypicality and its grammatical gender are marked in bold. Verb phrases used in the filler sentences were stereotypically gender-neutral. The sentence's and the question's translation to English are inserted in parentheses below.

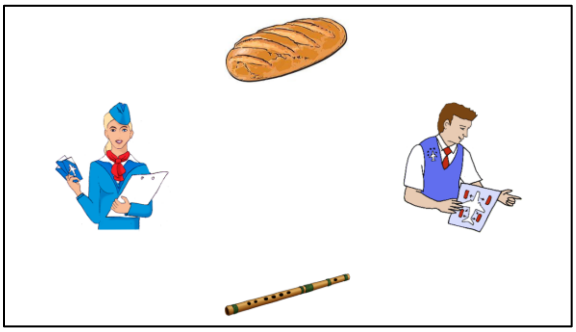
A Experiment 2 (Main Verbs - Humans) – Visual Stimuli: The Experimental Stimuli

Each of the visual scenes below represents the content of the two sentences within an experimental tuple. The number of the tuple the visual scene refers to is indicated below (see Table A29 for the full set of experimental sentences).

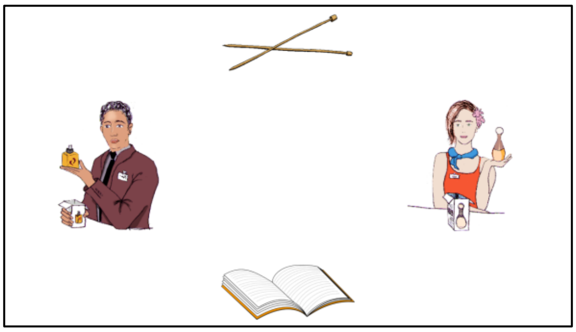


Tuple 1

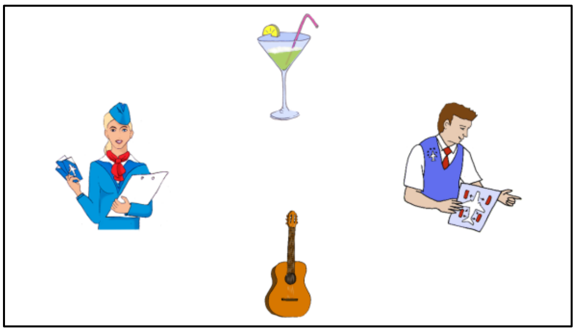
Tuple 2



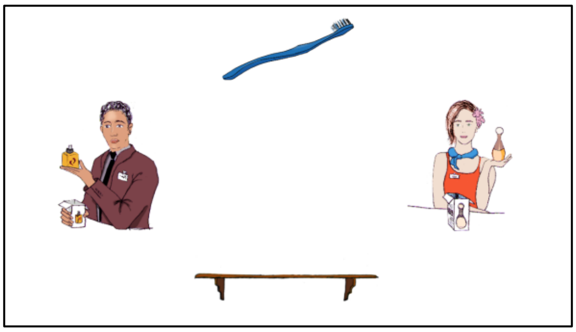
Tuple 3



Tuple 4



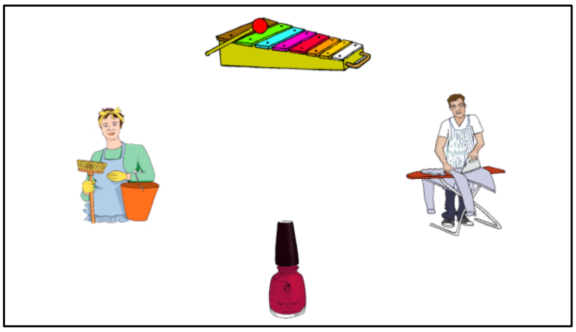
Tuple 5



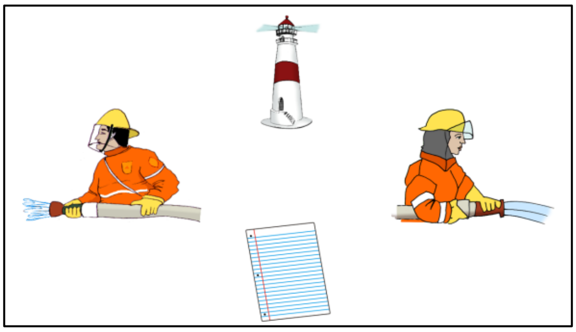
Tuple 6

(continued)

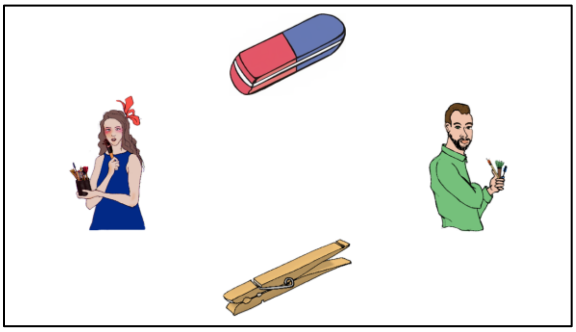
Each of the visual scenes below represents the content of the two sentences within an experimental tuple. The number of the tuple the visual scene refers to is indicated below (see Table A29 for the full set of experimental sentences).



Tuple 7

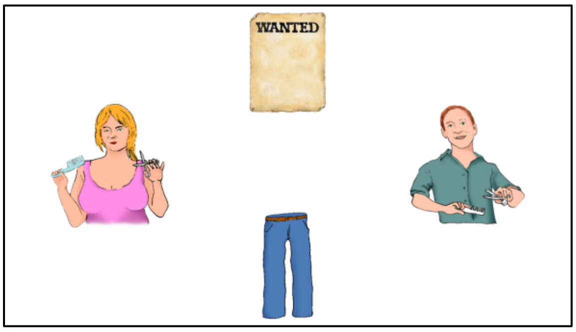


Tuple 8

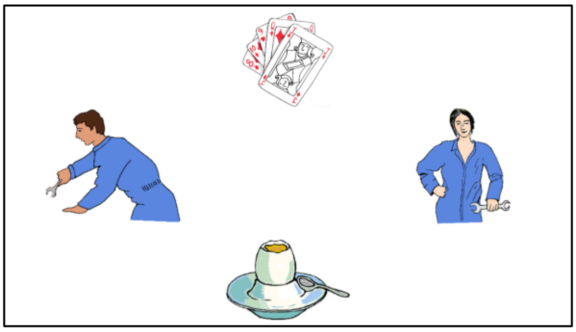


Tuple 9

Tuple 10



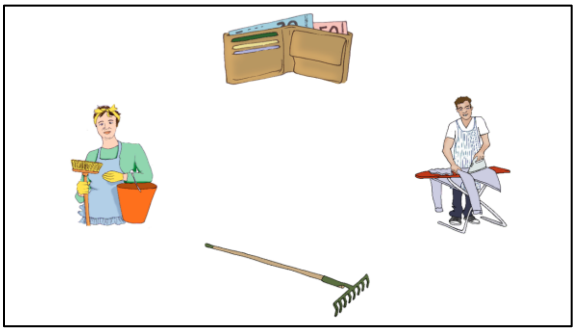
Tuple 11



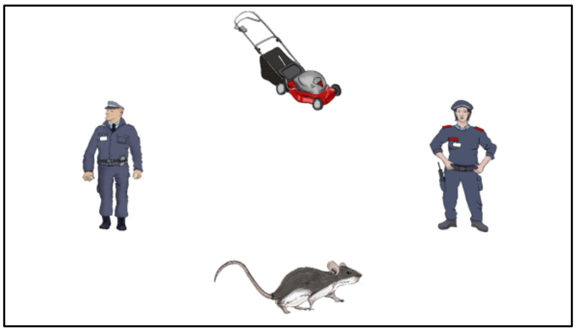
Tuple 12

(continued)

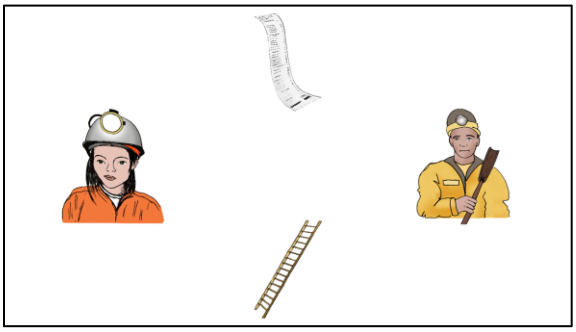
Each of the visual scenes below represents the content of the two sentences within an experimental tuple. The number of the tuple the visual scene refers to is indicated below (see Table A29 for the full set of experimental sentences).



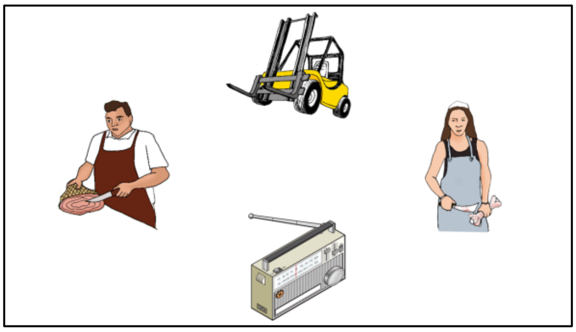
Tuple 13



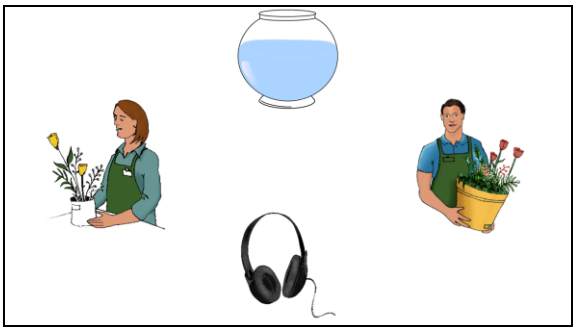
Tuple 14



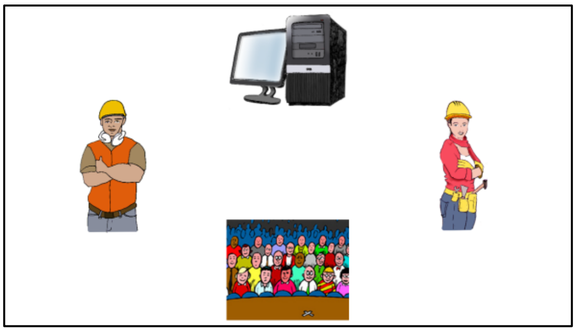
Tuple 15



Tuple 16



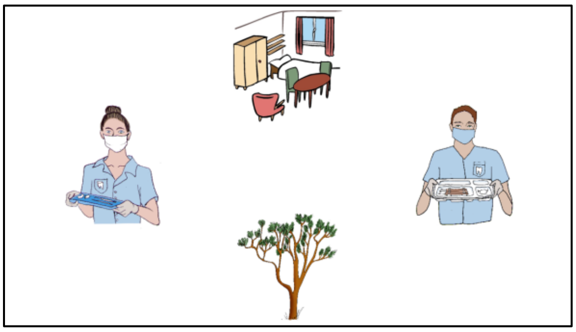
Tuple 17



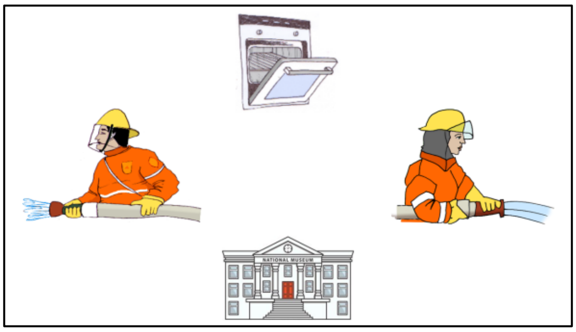
Tuple 18

(continued)

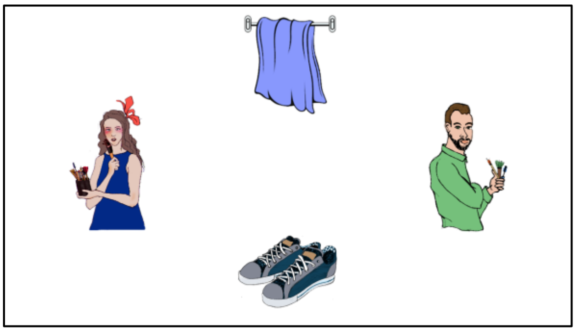
Each of the visual scenes below represents the content of the two sentences within an experimental tuple. The number of the tuple the visual scene refers to is indicated below (see Table A29 for the full set of experimental sentences).



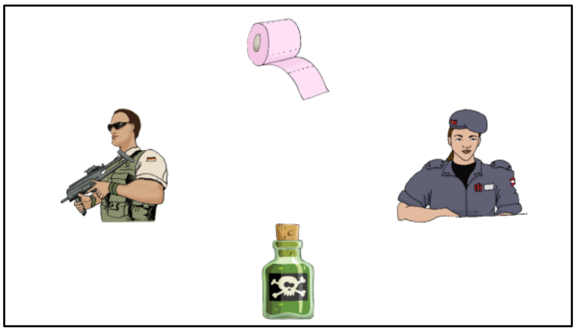
Tuple 19



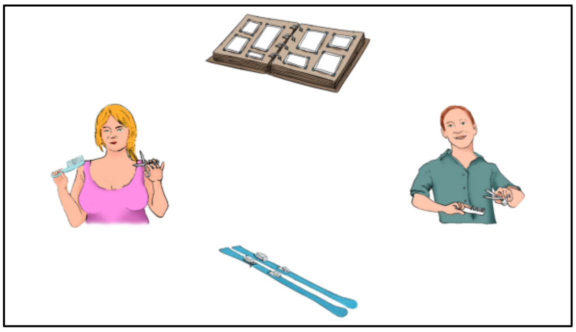
Tuple 20



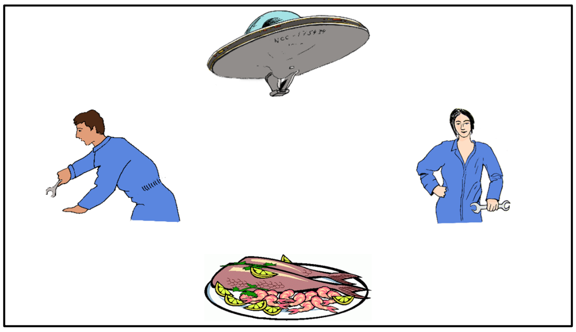
Tuple 21



Tuple 22



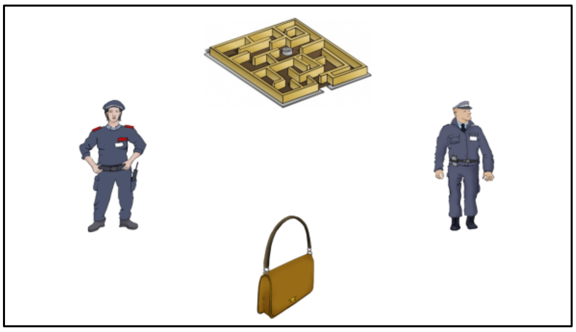
Tuple 23



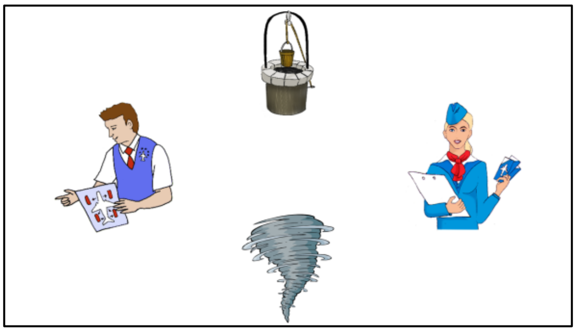
Tuple 24

(continued)

Each of the visual scenes below represents the content of the two sentences within an experimental tuple. The number of the tuple the visual scene refers to is indicated below (see Table A29 for the full set of experimental sentences).

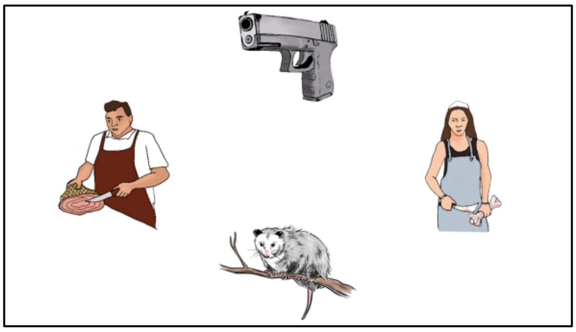


Tuple 25

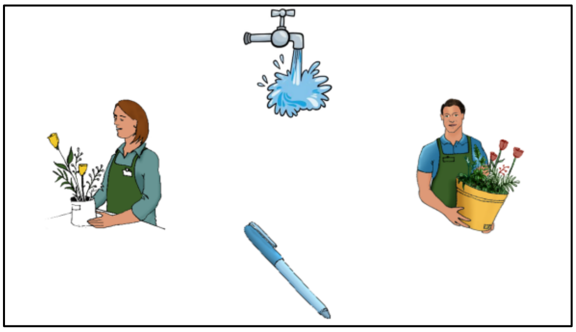


Tuple 26

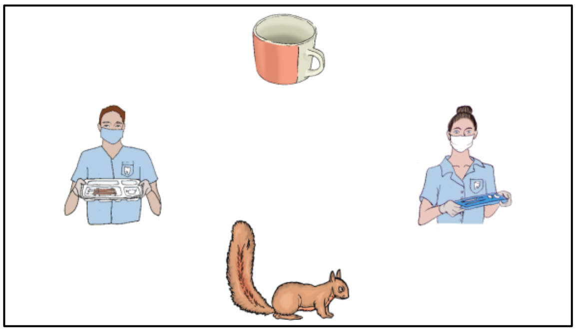
Tuple 27



Tuple 28



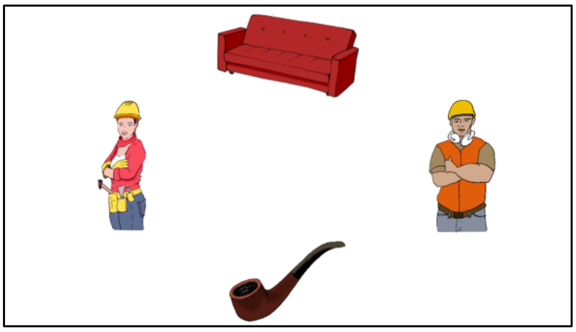
Tuple 29



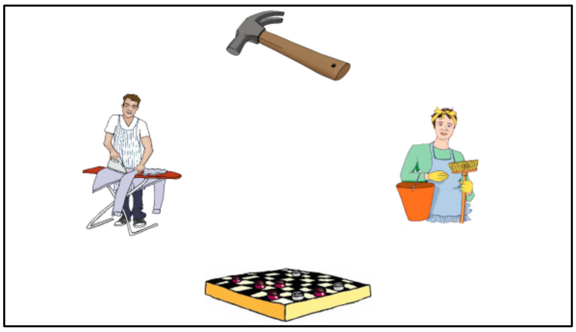
Tuple 30

(continued)

Each of the visual scenes below represents the content of the two sentences within an experimental tuple. The number of the tuple the visual scene refers to is indicated below (see Table A29 for the full set of experimental sentences).



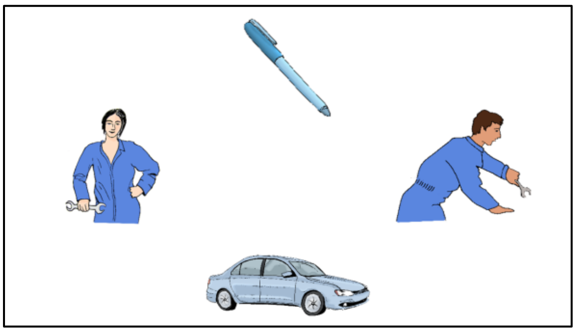
Tuple 31



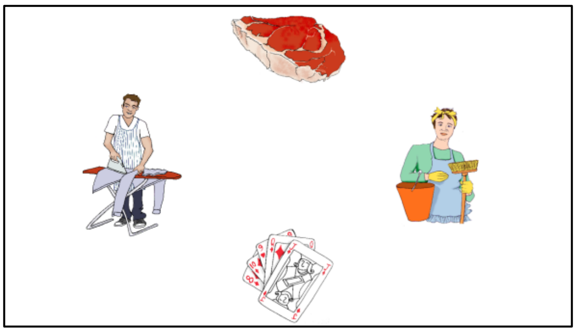
Tuple 32

A Experiment 2 (Main Verbs - Humans) – Visual Stimuli: The Fillers

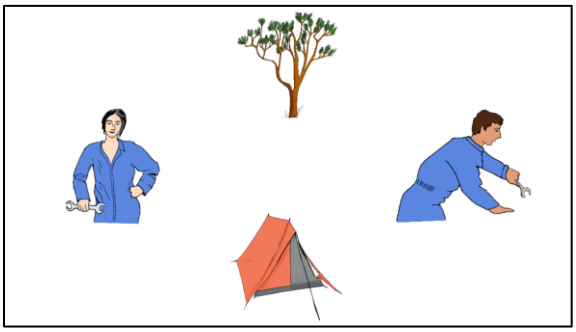
Each of the visual scenes below represents the content of a filler sentence. The number of the filler sentence a visual scene refers to is indicated below (see Table A30 for the full set of filler sentences).



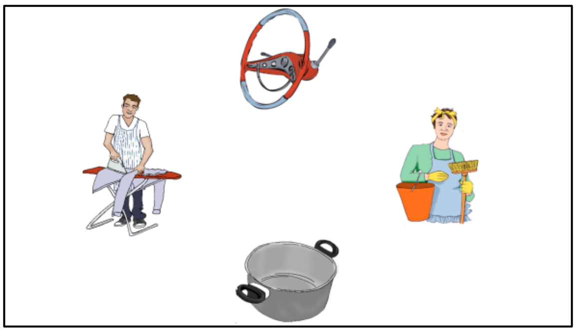
F1



F2



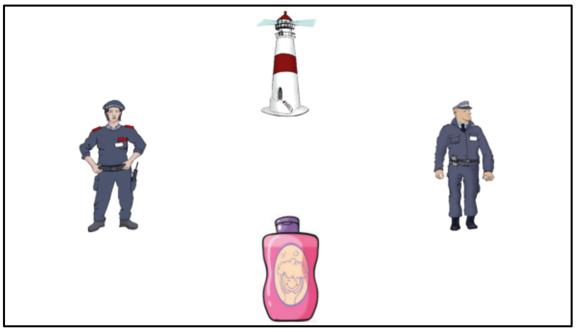
F3



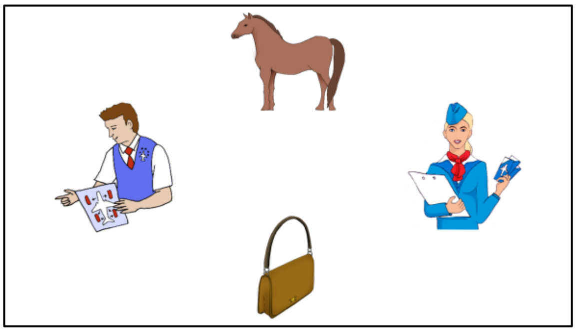
F4

(continued)

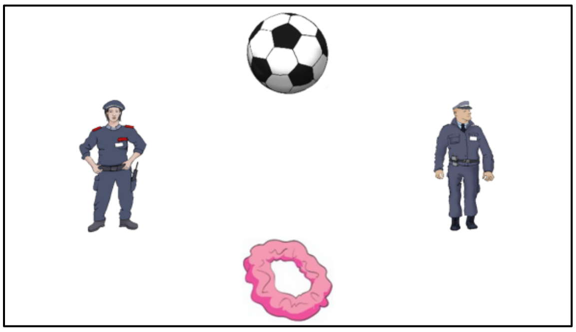
Each of the visual scenes below represents the content of a filler sentence. The number of the filler sentence a visual scene refers to is indicated below (see Table A30 for the full set of filler sentences) (continued).



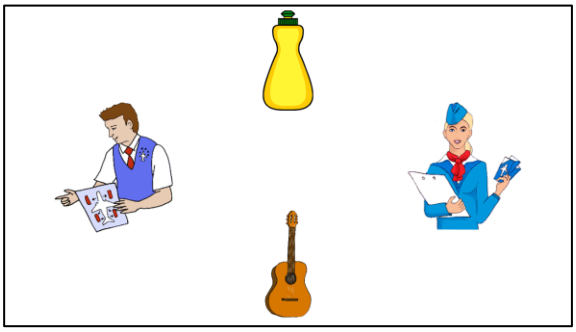
F5



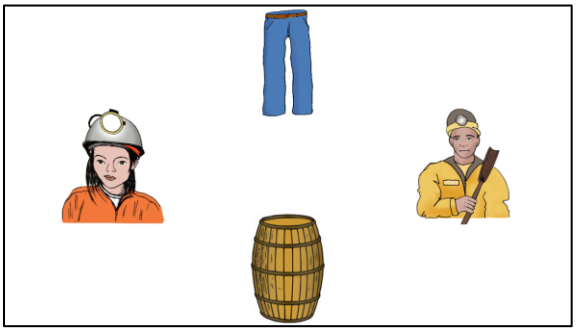
F6



F7



F8

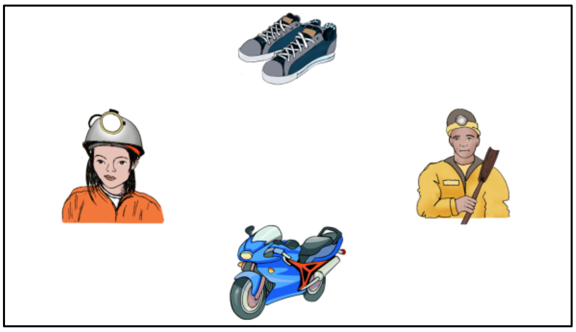


F9

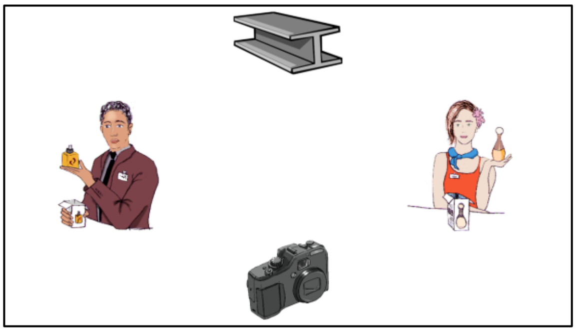
F10

(continued)

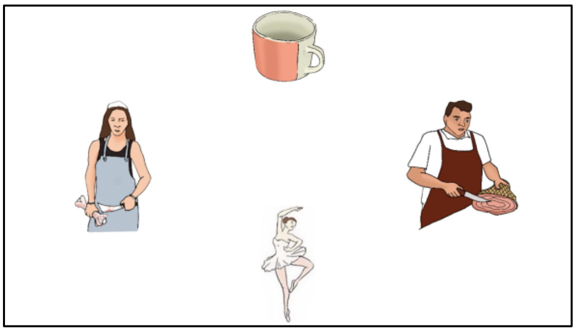
Each of the visual scenes below represents the content of a filler sentence. The number of the filler sentence a visual scene refers to is indicated below (see Table A30 for the full set of filler sentences) (continued).



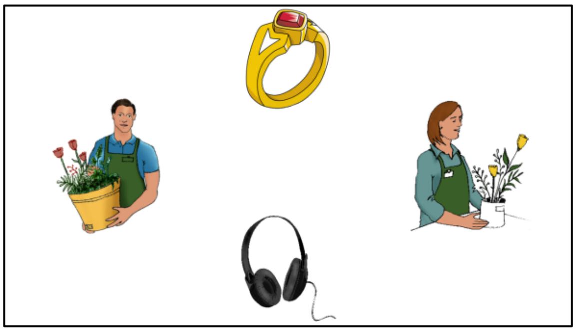
F11



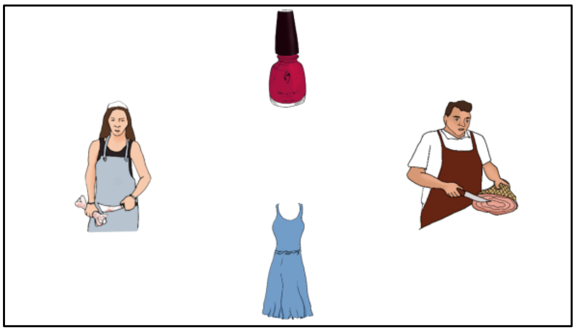
F12



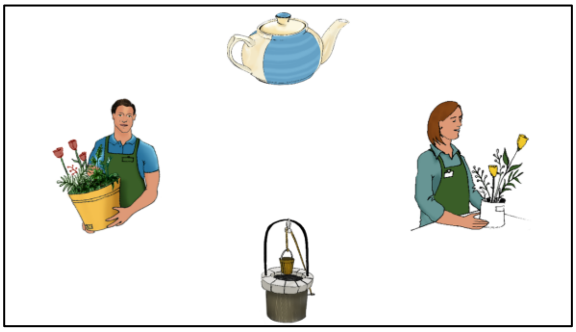
F13



F14



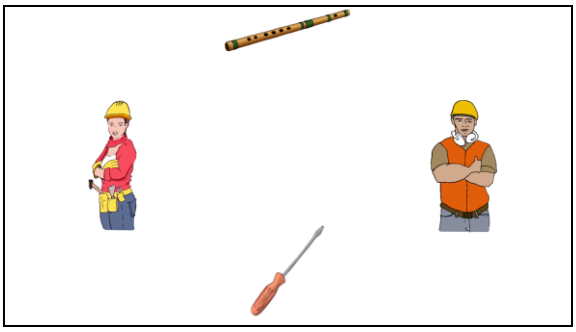
F15



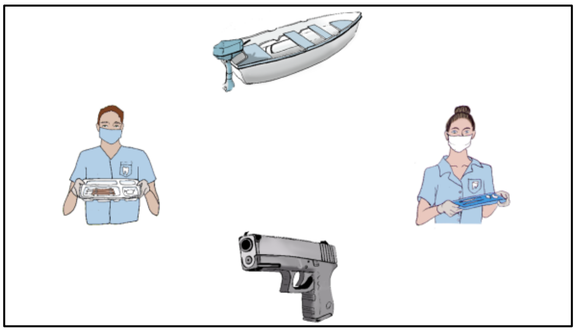
F16

(continued)

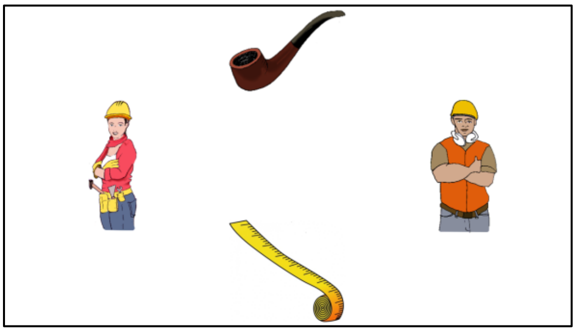
Each of the visual scenes below represents the content of a filler sentence. The number of the filler sentence a visual scene refers to is indicated below (see Table A30 for the full set of filler sentences) (continued).



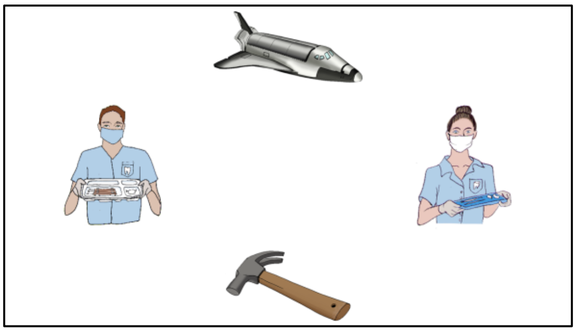
F17



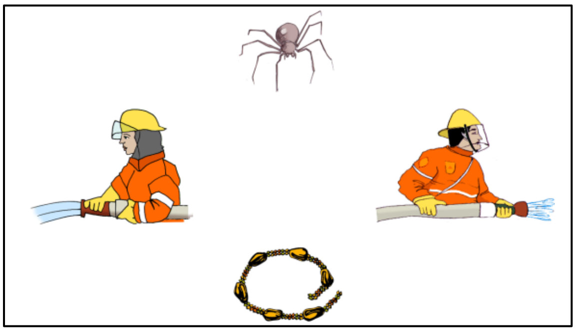
F18



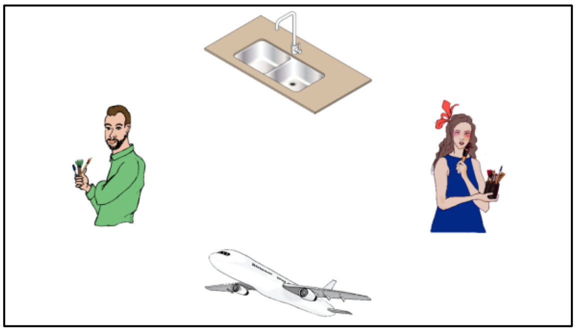
F19



F20



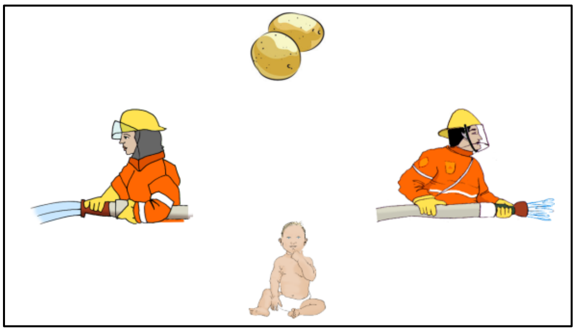
F21



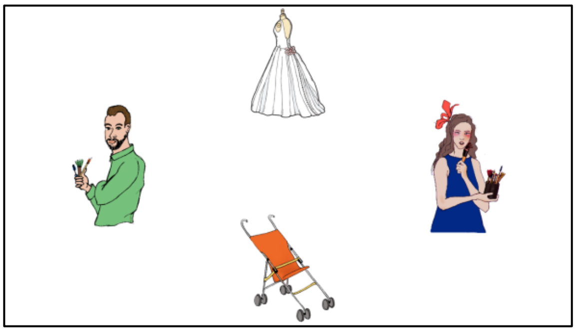
F22

(continued)

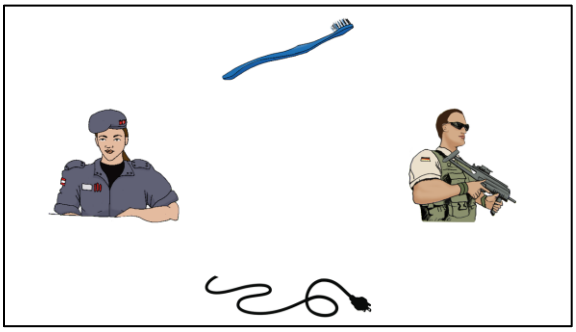
Each of the visual scenes below represents the content of a filler sentence. The number of the filler sentence a visual scene refers to is indicated below (see Table A30 for the full set of filler sentences) (continued).



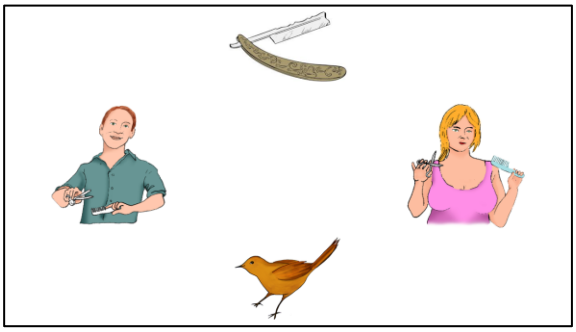
F23



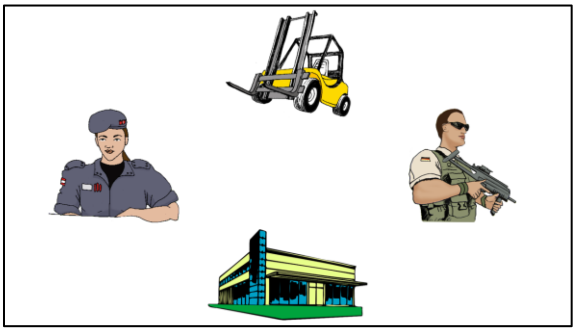
F24



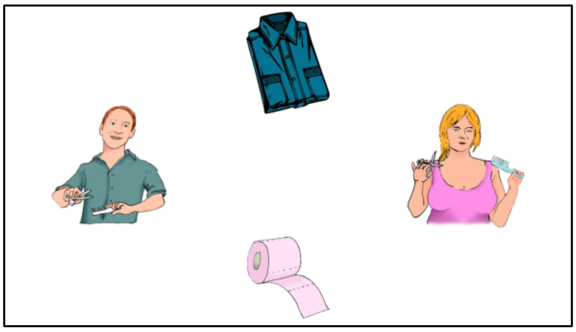
F25



F26



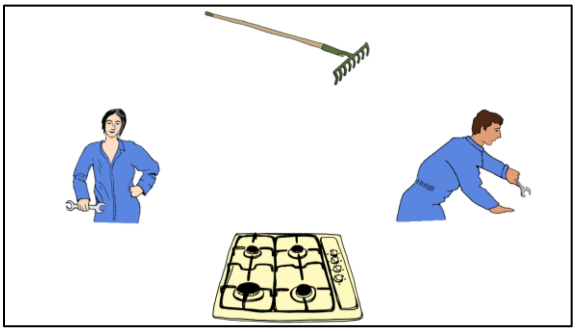
F27



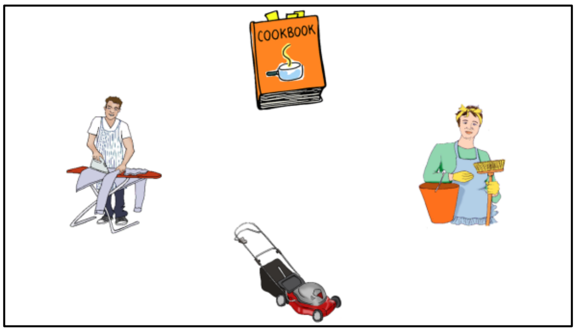
F28

(continued)

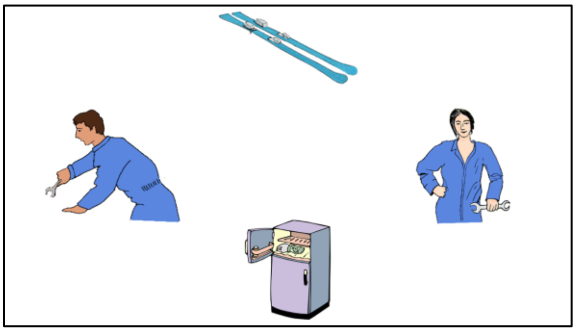
Each of the visual scenes below represents the content of a filler sentence. The number of the filler sentence a visual scene refers to is indicated below (see Table A30 for the full set of filler sentences) (continued).



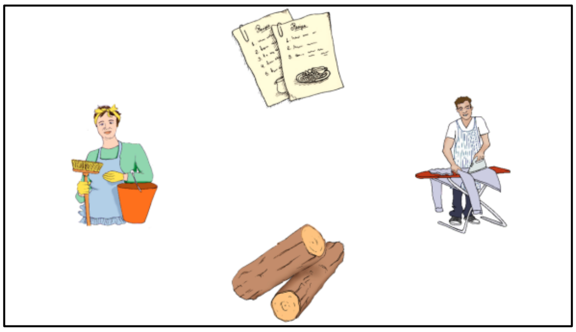
F29



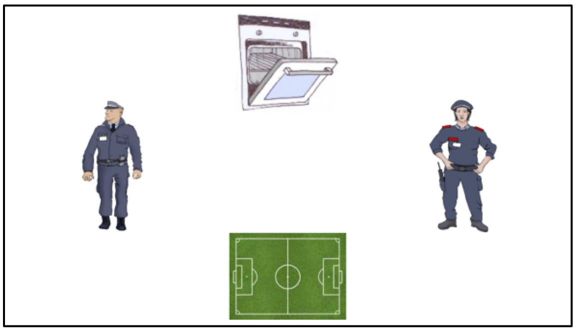
F30



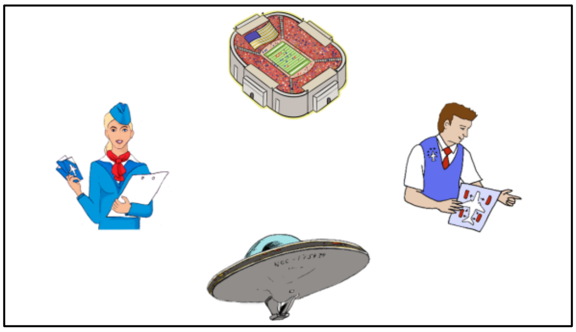
F31



F32



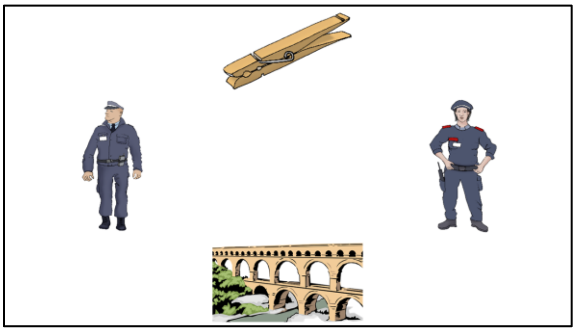
F33



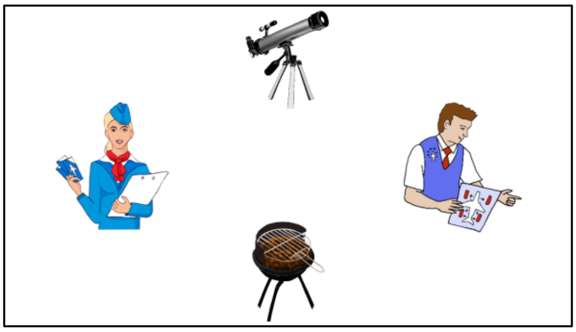
F34

(continued)

Each of the visual scenes below represents the content of a filler sentence. The number of the filler sentence a visual scene refers to is indicated below (see Table A30 for the full set of filler sentences) (continued).



F35

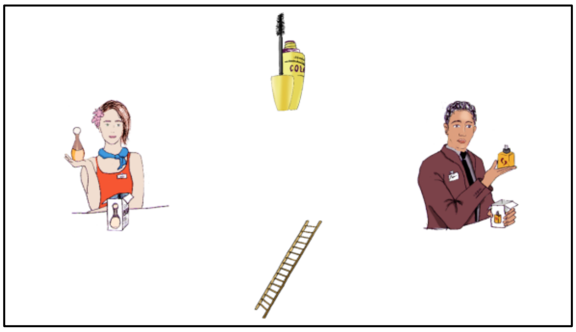


F36

F37

F38

F39



F40

A Pretest V – Pretest Instructions

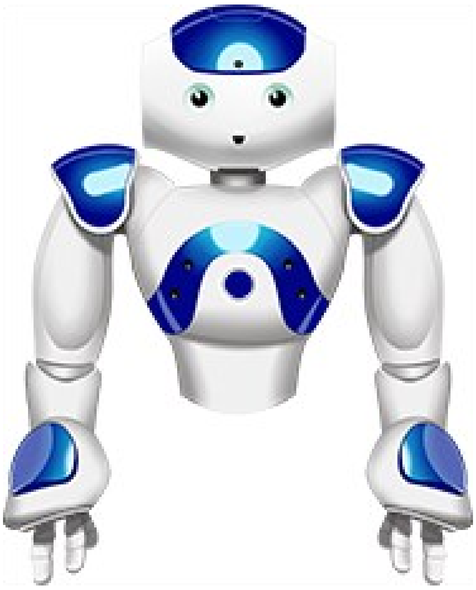
The pretest instructions in German (see Section 12.1). The translation to English can be found in parentheses below:

> Liebe Teilnehmerin, lieber Teilnehmer, wir führen eine Vorstudie zu einem neuen Roboter-Design durch. Bitte schauen Sie sich die Abbildung des Roboters genau an und beantworten Sie anschließend die folgenden Fragen. Uns interessiert Ihr persönlicher Eindruck. Es gibt daher keine richtigen oder falschen Antworten. Die Bearbeitung des Fragebogens wird nicht länger als 10 Minuten dauern. Vielen Dank für Ihre Teilnahme! Beginnen Sie nun mit der Bearbeitung des Fragebogens auf der Rückseite.

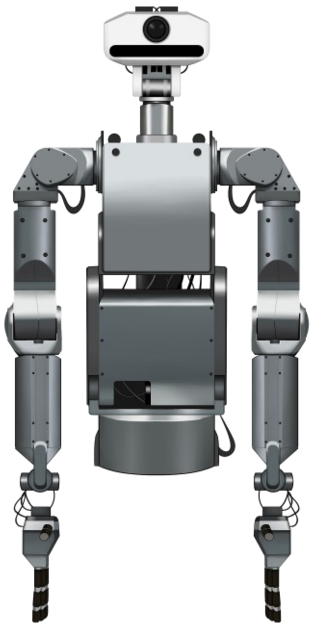
[Dear participant, we conduct a pretest on a new robot design. Please look carefully at the depicted robot before answering the following questions. We are interested in your personal opinion. Therefore, there are no right or wrong answers. Completing the questionnaire will take no more than 10 minutes. Thanks for your participation! Please start now answering the questions on the back side.]

A Pretest V – Material: Identification of 'Typical' Robots

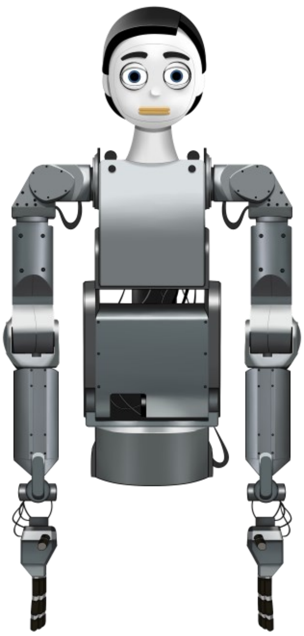
These drawings of existing robots were presented in Pretest V (see Section 12.1). The robot's name and the company's/university's name that developed the robot (head or/and body) are listed in parentheses.



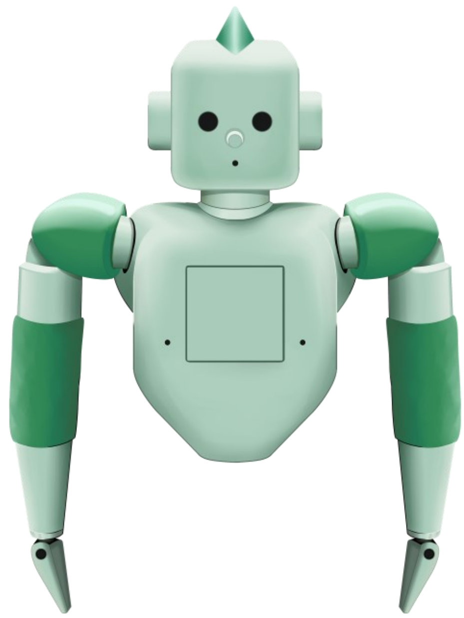
1. NAO
(*SoftBank Robotics*)



2. *Meka* body with
Meka sensor head (*Google X*)



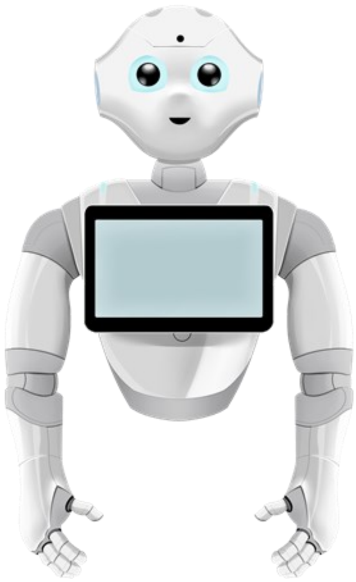
3. *FloKa: Meka* body (*Google X*) with
FloBi head (*Bielefeld University*)



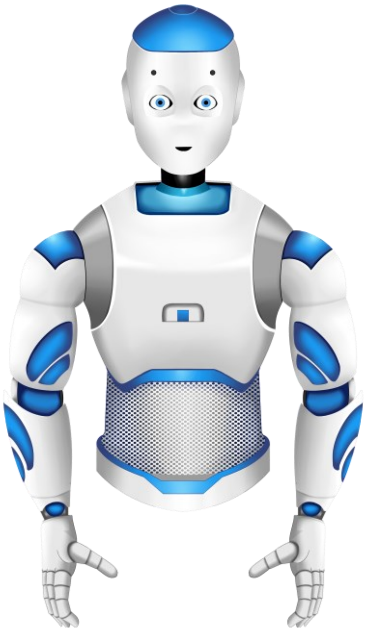
4. Ri-Man (*Riken Bio-Mimetic Control, Research Center Nagoya*)

(continued)

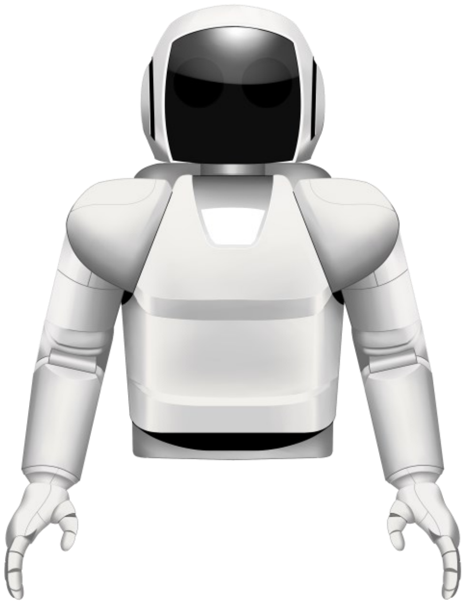
These drawings of existing robots were presented in Pretest V (see Section 12.1). The robot's name and the company's/university's name that developed the robot (head or/and body) are listed in parentheses (continued).



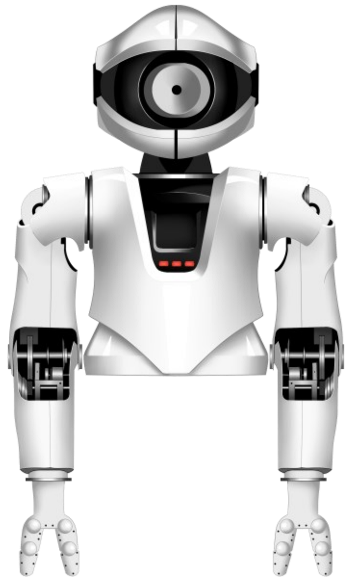
5. *Pepper*
(*SoftBank Robotics*)



6. *ROMEO*
(*SoftBank Robotics*)



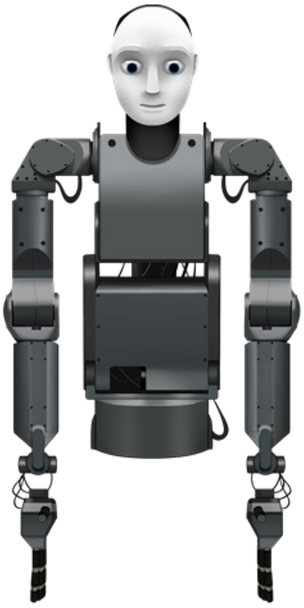
7. *ASIMO*
(*Honda*)



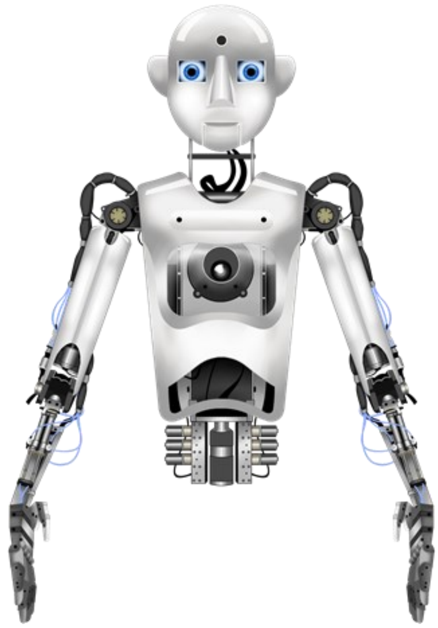
8. *Myon*
(*Beuth Hochschule für Technik Berlin*)

(continued)

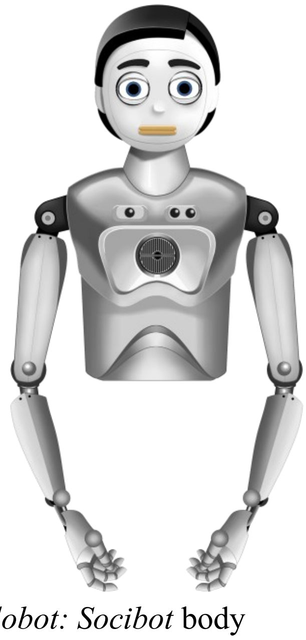
These drawings of existing robots were presented in Pretest V (see Section 12.1). The robot's name and the company's/university's name that developed the robot (head or/and body) are listed in parentheses (continued).



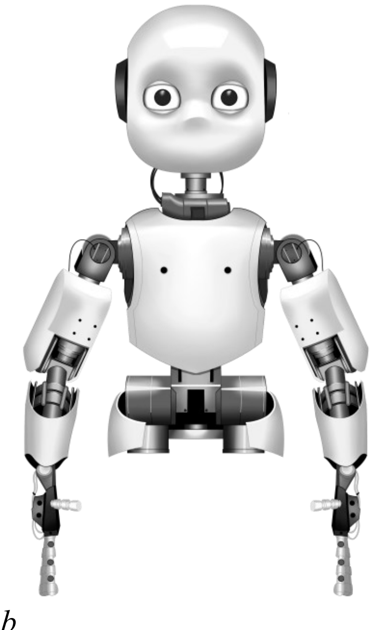
9. *Folkwang: Meka* body (*Google* X) with *Folkwang* head (*Folkwang University*)



10. *RoboThespian* (*Engineered Arts*)



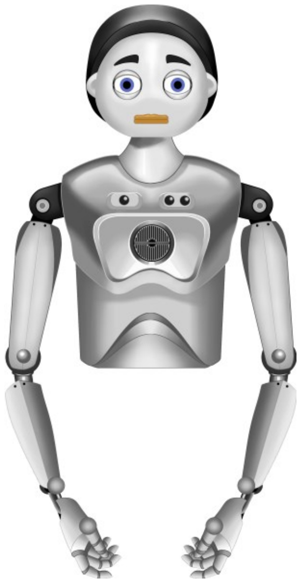
11. *SociFlobot: Socibot* body (*Engineered Arts*) with *FloBi* head (*Bielefeld University*)



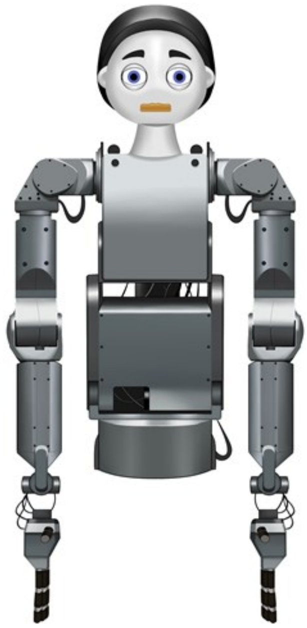
12. *iCub* (*Italian Institute of Technology, IIT*)

(continued)

These drawings of existing robots were presented in Pretest V (see Section 12.1). The robot's name and the company's/university's name that developed the robot (head or/and body) are listed in parentheses (continued).



13. *SociFlobot2: Socibot* body (*Engineering Arts*) with redesigned *FloBi* head (*Bielefeld University*)



14. *FloKa2: Meka* body (*Google* X) with redesigned *FloBi* head (*Bielefeld University*)

A Pretest V – Results: Participants' Perceptions of 'Typical' Robot Characteristics and of Robot Gender

In Pretest V, participants were presented with one of 14 robot drawings which they had to judge in terms of *recognizability* and *typicality* using 7-point Likert scales. Participants' ratings of the robot drawings' typicality and recognizability were reported in Section 12.1, Table 12.1. Participants' judgements of what characteristics made the robot drawings appear typical or untypical for a robot, their perceptions of the robot drawings as male or female, and their ratings of the gender-stereotypical profession the robots could represent are reported in the following.

More specifically, using open-response formats, participants were asked to indicate what characteristics of the robot drawing they had to judge they perceived as *typical* or as *untypical* for a robot and what features would make the robot drawing appear as more typical. Likewise, using open-response formats, participants were asked to specify what characteristics made the robot drawing appear male or female. Finally, the gender-stereotypical professions were listed. Using 7-point Likert scales, participants had to indicate to what extent the robot they had to judge could portray each of the listed professions (see also Section 12.1).

A Pretest V – Results: Characteristics of the Robot Drawings That Were Considered Typical for a Robot

Joints, hands, a torso, facial cues and hair parts, a helmet, metal equipment, a camera eye, metal frames consisting of single parts, humanlike appearance, metal hinges, metal surface (like on tv), screws, metal blocks, chrome, cubic form, tools (e.g., arms) and sensors (e.g., eyes), cables, rigid appearance, made of single parts, technical elements, mechanic arms, grippers, head mounted camera, display, holistic humanlike body, cubic, mimicking humans, expressionless eyes, 'cold' colors (e.g., blue, silver, and white surface or lights), metal 'clothes', android design, visible skeleton (like 'C3PO'), stylistic form (like in sci-fi movies), many single joints, mechanic features, metal gutter, hinges, visible technical construction, basically humanlike but cubic, stiff, no face but cameras speakers and sensors instead, no emotional expression, mechanic movements, inorganic surface, metal, wires, cables, electronic features, geometric form, humanlike features but mechanic fragments, symmetric and futuristic design, and expressionless appearance (like in 'iRobot') were deemed as typical for a robot.

A Pretest V – Results: Characteristics of the Robot Drawings That Were Considered Untypical for a Robot

A friendly face, a comic-like appearance, single fingers, lips, eye brows, and humanlike cues that differed from humans (e.g., *Myon* robot has one eye, *NAO* robot has four fingers) were deemed as untypical for a robot.

A Pretest V – Results: Participants' Suggestions How to Make the Robot Drawings Seem More Typical for a Robot

In line with their indications of the robots' typicality (see Table 12.1), participants explained that the robots already looked like typical exemplars of the category 'robot'. To make them appear even more typical, the robots should look more cubic, technical (e.g., with gear wheels), 'efficient', dangerous (e.g., with red eyes), and should have a 'cold' color. They should not have a face or detailed cues, such as eyes, eye brows (e.g., like *FloBi* robot), or fingers (e.g., like *NAO* robot). If robots had such humanlike cues, these should be realistic. To illustrate, unlike *Myon* or NAO of which one has only one eye, while the other has only four fingers, a robot should have two eyes and five fingers like humans do.

A Pretest V – Results: Characteristics Participants Listed as Male and as Female for Robots

Using an open-response format, participants were asked to indicate what features of the robot drawing they perceived as male or as female. In line with existing literature (see Bernotat et al.,2017; 2021; Eyssel & Hegel, 2012), participants reported that a robot's shoulder width, waist form, hair length, and lip color were decisive for the attribution of robot gender. According to participants' judgements, male robots should have broad shoulders and short hair, while female robots should have a narrow waist, long hair, and red lips.

A Pretest V – Results: Participants' Ratings of the Robots' Suitability to Portray the Gender-Stereotypical Professions

To investigate to what extent participants judged the robots as suitable to represent one of the professions, one-sample *t*-tests were performed against the scale midpoint of 4 (see Table A31).

Table A31

Participants' Mean Ratings (Standard Deviations in Parentheses) of the Robots' Suitability to Portray Gender-Stereotypical Professions Tested Against the Scale Midpoint of 4

Profession	1. ASIMO	2. Floka	3. Floka 2	4. Folkwang
Bouncer	3.70 (1.98)	3.24 (1.90)†	2.39 (1.70)***	4.21 (2.15)
Butcher	3.19 (1.96)*	3.88 (1.69)	3.41 (1.84)	4.25 (1.89)
Construction Worker	3.78 (2.31)	3.64 (1.63)	3.39 (1.83)	4.13 (1.85)
Cosmetician	1.44 (0.85)***	3.04 (1.93)*	2.83 (1.40)**	1.79 (1.25)***
Dental Assistant	3.04 (2.07)*	4.12 (1.36)	4.00 (1.85)	3.33 (1.99)
Firefighter	3.96 (2.12)	3.40 (1.63)†	2.55 (1.63)***	3.04 (2.05)*
Florist	2.04 (1.22)***	3.92 (1.53)	3.00 (1.57)**	2.75 (1.70)**
Gold Digger	3.56 (2.01)	3.24 (1.76)*	4.00 (2.14)	3.67 (2.30)
Hair Dresser	1.96 (1.09)***	3.40 (1.73)†	3.17 (1.61)*	1.92 (1.35)***
Housewife	3.50 (2.20)	4.96 (1.81)*	4.26 (1.79)	3.83 (1.93)
Mechatronics Engineer	5.04 (2.14)*	5.36 (1.11)***	4.73 (1.86)†	4.71 (2.07)
Perfumery Shop Assistant	1.89 (1.58)***	2.56 (1.45)***	2.17 (1.37)***	1.83 (1.17)***
Soldier	4.85 (2.07)*	3.64 (1.89)	3.09 (2.00)*	4.71 (2.20)
Stewardess	2.81 (1.67)**	4.44 (1.87)	3.09 (1.66)*	3.33 (1.93)

Note. ***$p < .001$, **$p < .01$, *$p < .05$, †$p < .10$.

(continued)

Table A31

Participants' Mean Ratings (Standard Deviations in Parentheses) of the Robots' Suitability to Portray Gender-Stereotypical Professions Tested Against the Scale Midpoint of 4 (continued)

Profession	5. iCub	6. Meka	7. Myon	8. NAO
Bouncer	2.81 (1.83)**	4.04 (2.06)	3.40 (1.89)	3.07 (2.21)*
Butcher	2.71 (1.90)**	4.00 (1.77)	3.36 (2.08)	3.00 (1.72)**
Construction Worker	3.43 (1.86)	4.04 (1.90)	3.92 (1.63)	3.07 (1.86)*
Cosmetician	2.43 (1.81)**	1.92 (1.14)***	2.16 (1.41)***	2.54 (1.48)***
Dental Assistant	3.38 (1.99)	3.79 (2.00)	3.96 (1.93)	4.48 (1.85)
Firefighter	2.62 (1.40)***	3.21 (1.74)*	3.04 (1.84)*	4.07 (2.24)
Florist	3.05 (1.53)*	2.42 (1.59)***	4.00 (1.92)	3.71 (1.78)
Gold Digger	3.33 (1.98)	3.92 (1.93)	3.60 (2.06)	2.89 (1.87)**
Hair Dresser	2.71 (1.74)**	2.54 (1.50)***	2.68 (1.82)**	2.75 (1.56)***
Housewife	4.00 (2.17)	3.92 (1.79)	4.40 (2.02)	4.61 (1.73)†
Mechatronics Engineer	3.71 (1.90)	5.35 (1.30)***	4.84 (1.97)*	4.18 (2.04)
Perfumery Shop Assistant	2.62 (1.80)**	1.83 (1.40)***	2.40 (1.89)***	2.71 (1.51)***
Soldier	2.62 (1.53)**	4.63 (2.00)	4.12 (1.97)	3.61 (2.38)
Stewardess	3.81 (1.81)	3.79 (1.96)	4.36 (1.98)	4.68 (1.77)†

Note. ***$p < .001$, **$p < .01$, *$p < .05$, †$p < .10$.

(continued)

Table A31

Participants' Mean Ratings (Standard Deviations in Parentheses) of the Robots' Suitability to Portray Gender-Stereotypical Professions Tested Against the Scale Midpoint of 4 (continued)

Profession	9. Pepper	10. Ri-Man	11. RoboThespian	12. Romeo
Bouncer	2.03 (1.48)***	3.24 (2.09)†	3.35 (2.11)	2.85 (1.98)**
Butcher	2.55 (1.86)***	3.36 (1.63)†	4.05 (1.91)	2.63 (1.84)**
Construction Worker	1.93 (1.53)***	4.20 (1.87)	3.80 (2.19)	2.48 (1.81)***
Cosmetician	3.14 (2.12)*	2.24 (1.89)***	2.75 (1.77)**	2.96 (2.16)*
Dental Assistant	4.45 (1.90)	4.08 (1.94)	4.15 (2.18)	3.85 (2.55)
Firefighter	2.68 (1.95)**	3.16 (1.82)*	3.55 (2.01)	3.11 (1.87)*
Florist	3.17 (1.65)*	3.16 (1.89)*	3.25 (1.77)†	3.70 (2.05)
Gold Digger	2.31 (2.12)***	3.58 (2.39)	3.55 (2.16)	2.41 (1.45)***
Hair Dresser	3.41 (2.13)	2.28 (1.59)***	3.00 (2.13)*	2.93 (1.98)**
Housewife	4.69 (1.91)†	4.40 (1.76)	4.50 (2.14)	4.56 (1.99)
Mechatronics Engineer	4.17 (2.04)	4.72 (2.07)†	4.75 (1.68)†	4.85 (1.75)*
Perfumery Shop Assistant	2.55 (1.82)***	2.28 (1.72)***	2.50 (1.64)**	2.41 (1.65)***
Soldier	3.07 (2.30)*	3.68 (2.30)	3.05 (2.06)†	3.56 (2.19)
Stewardess	4.96 (1.92)*	3.88 (2.03)	4.20 (1.94)	4.70 (2.22)

Note. ***$p < .001$, **$p < .01$, *$p < .05$, $^{\dagger}p < .10$.

(continued)

Table A31

Participants' Mean Ratings (Standard Deviations in Parentheses) of the Robots' Suitability to Portray Gender-Stereotypical Professions Tested Against the Scale Midpoint of 4 (continued)

Profession	13. SociFlobot	14. SociFlobot 2
Bouncer	3.86 (2.10)	3.30 (1.98)
Butcher	3.67 (2.24)	3.15 (1.66)*
Construction Worker	3.29 (1.88)†	3.40 (1.10)*
Cosmetician	3.19 (1.69)*	3.40 (1.67)
Dental Assistant	3.81 (1.81)	4.50 (1.73)
Firefighter	3.19 (1.91)†	3.20 (1.74)†
Florist	4.24 (1.58)	3.85 (2.06)
Gold Digger	3.33 (2.13)	3.40 (1.67)
Hair Dresser	4.14 (1.98)	3.65 (1.98)
Housewife	4.52 (1.94)	4.79 (1.48)*
Mechatronics Engineer	4.48 (2.04)	4.70 (1.17)*
Perfumery Shop Assistant	3.10 (2.05)†	3.35 (1.90)
Soldier	4.24 (2.30)	3.16 (2.01)†
Stewardess	4.33 (1.62)	4.42 (2.17)

Note. ***$p < .001$, **$p < .01$, *$p < .05$, $^{\dagger}p < .10$.

A Pretest VI – Pretest Instructions

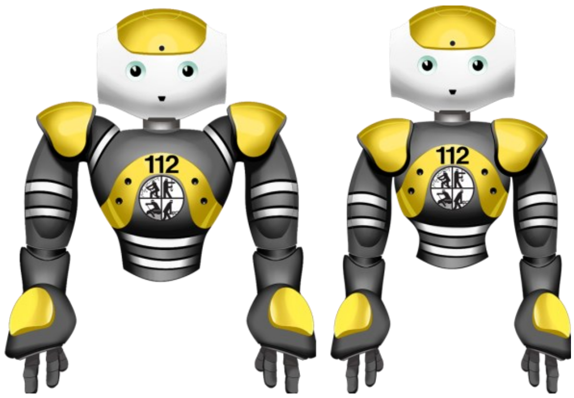
The pretest instructions in German (see Section 12.2). The translation to English can be found in parentheses below:

> Liebe Teilnehmerin, lieber Teilnehmer, vielen Dank, dass du dir Zeit nimmst, an dieser Vorstudie teilzunehmen. Die Bearbeitung des Fragebogens wird nicht länger als 15 Minuten dauern.
>
> Im Folgenden wirst du Abbildungen verschiedener Roboter-Paare sehen. Diese Roboter-Paare stellen jeweils verschiedene Berufe dar. Bitte schau dir die Abbildungen genau an. Schätze dann ein, welchen Beruf das jeweilige Roboter-Paar deiner Ansicht nach darstellt. Bitte antworte so spontan wie möglich. Es gibt hierbei keine richtigen oder falschen Antworten. Uns interessiert deine persönliche Einschätzung. Bei Bedarf erhältst du anschließend 0.5 VP-Stunden für deine Teilnahme.
>
> Deine Daten werden streng vertraulich behandelt und nur in anonymisierter Form für wissenschaftliche Zwecke verwendet. Es sind zu keinem Zeitpunkt Rückschlüsse auf deine Person möglich. Die Teilnahme ist freiwillig und kann jederzeit ohne Angaben von Gründen beendet werden, ohne dass dir daraus Nachteile entstehen. In diesem Fall würden deine Daten vollständig gelöscht werden. Vielen Dank für deine Teilnahme! Mit einem Klick auf „weiter“ stimmst du der Nutzung deiner anonymisierten Daten zu und kannst mit der Bearbeitung des Fragebogens beginnen.

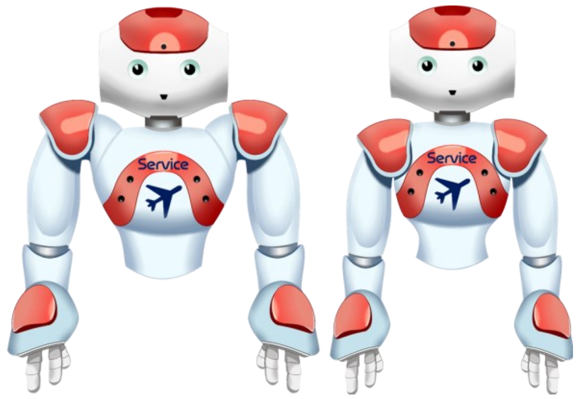
[Dear participant, thank you for taking time to support this pretest. The completion of the questionnaire will take no more than 15 minutes. In the following, you will be presented pictures of robot pairs. Each robot pair represents a certain profession. Please look at the robot pairs carefully. Then, indicate which profession the robots might represent. Please respond as spontaneously as possible. There are no right or wrong answers. We are just interested in your personal guesses. You receive 0.5 course credits for participation if needed. Your data will be kept confidential and only used in an anonymized form. Therefore, no conclusions can be drawn on your identity. Your participation is voluntary and can be ended every time without giving any reasons. In this case, no consequences will follow and your data will be entirely deleted. Thanks for your participation! By clicking the “next”-button, you give consent your anonymized data to be used and you can start with the completion of the questionnaire.]

A Pretest VI – Material: Evaluation of the Robot Drawings' Gender and Profession

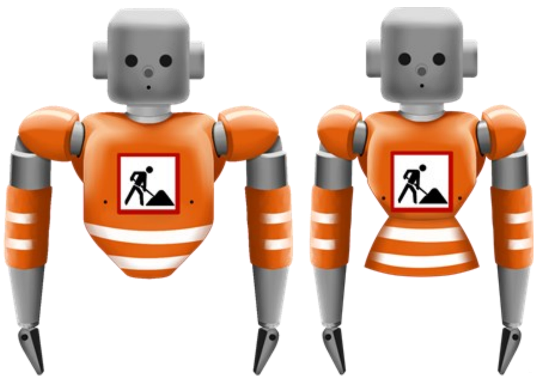
These drawings of the robot pairs were presented in Pretest VI (see Section 12.2). The male robot is depicted left to the female one. The robots' name, the profession they represent, and, in parentheses, the profession's gender-stereotypicality are listed below.



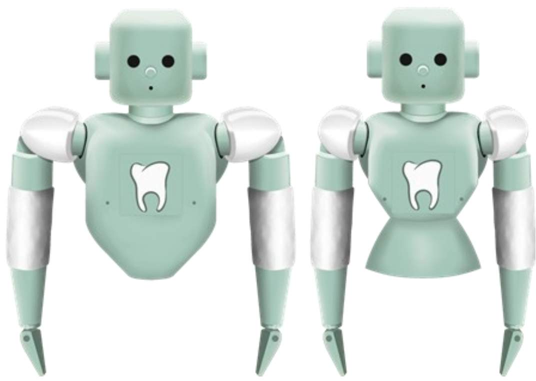
1. *Pepper*:
Firefighter (male)



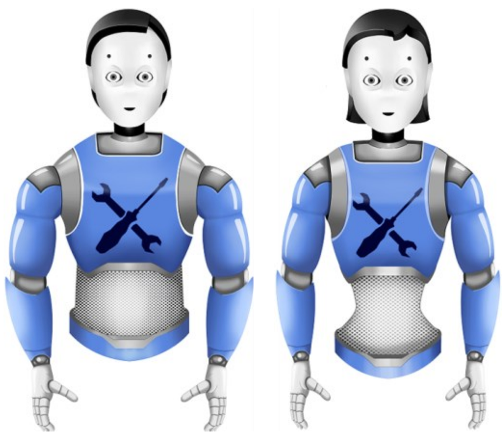
2. *Pepper*:
Stewardess (female)



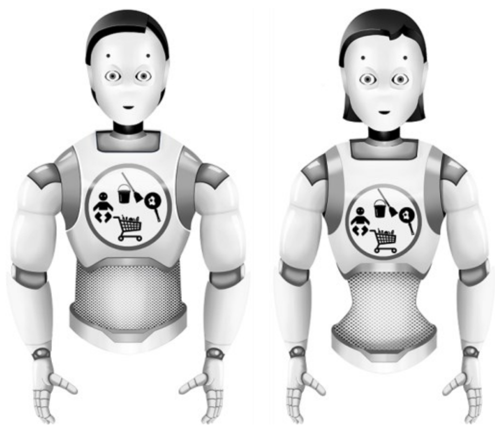
3. *Ri-Man:*
Construction Worker (male)



4. *Ri-Man:*
Dental Assistant (female)



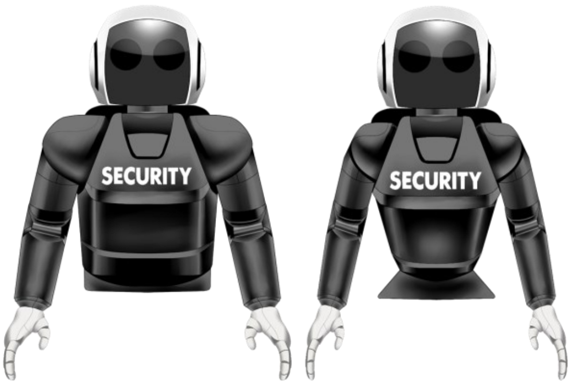
5. *Romeo:*
Mechatronics Engineer (male)



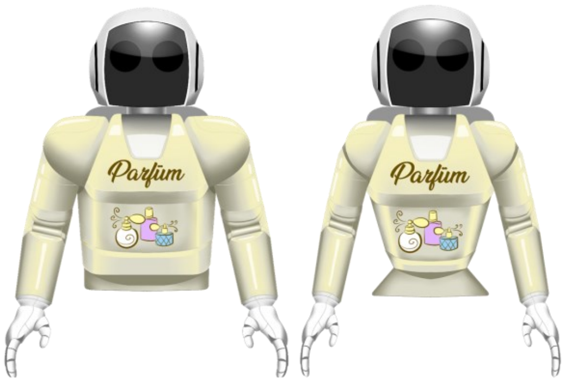
6. *Romeo:*
Housewife (female)

(continued)

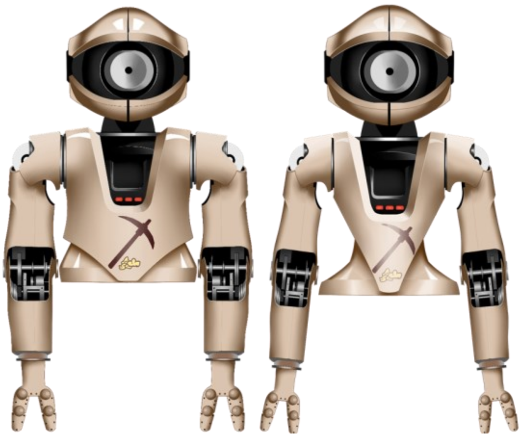
These drawings of the robot pairs were presented in Pretest VI (see Section 12.2). The male robot is depicted left to the female one. The robots' name, the profession they represent, and, in parentheses, the profession's gender-stereotypicality are listed below (continued).



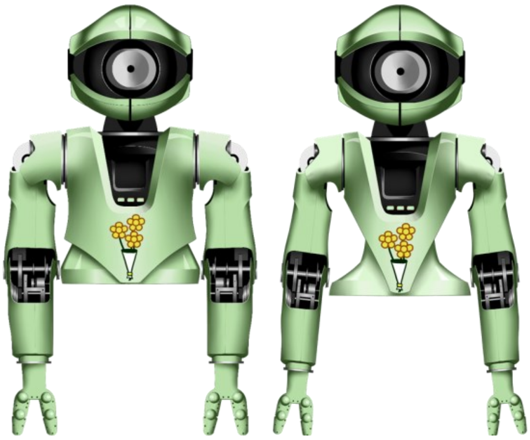
7. *ASIMO*:
Bouncer (male)



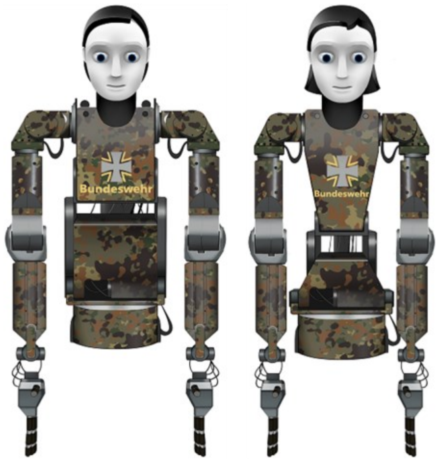
8. *ASIMO*:
Perfumery Shop Assistant (female)



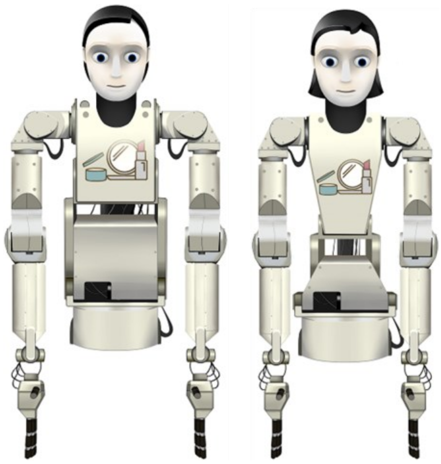
9. *Myon:*
Gold Digger (male)



10. *Myon:*
Florist (female)



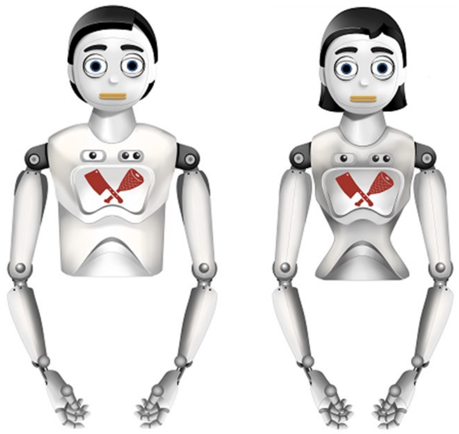
11. *Folkwang:*
Soldier (male)



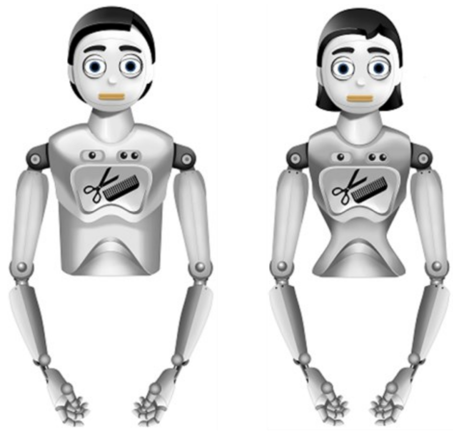
12. *Folkwang:*
Cosmetician (female)

(continued)

These drawings of the robot pairs were presented in Pretest VI (see Section 12.2). The male robot is depicted left to the female one. The robots' name, the profession they represent, and, in parentheses, the profession's gender-stereotypicality are listed below (continued).



13. *SociFlobot*:
Butcher (male)

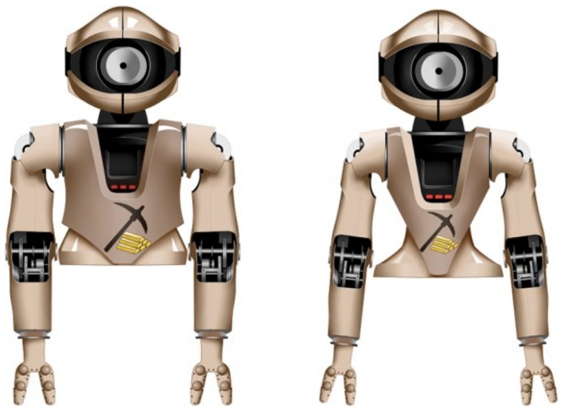


14. *SociFlobot*:
Hair Dresser (female)

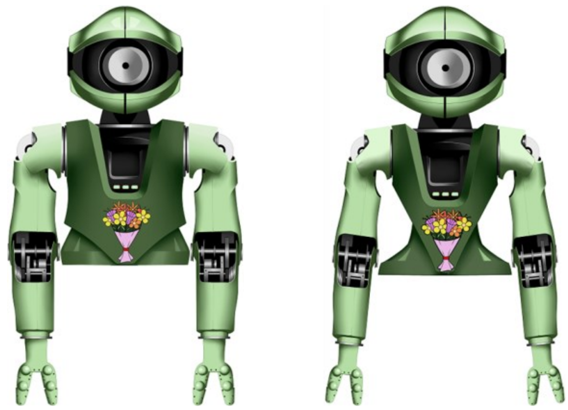
A Experiment 3 (Adverbs – Robots) – Visual Stimuli: Final Robot Drawings

Most of the robots that were displayed in Experiment 3 were presented as in Pretest VI (see Section 12.2, see robot drawings pp. 286). Only the robots that were portrayed as fire-fighters, gold diggers, and florists were edited based on the results of Pretest VI. The edited final drawings are displayed below. The robots' name, the profession they represented, and, in parentheses, the profession's gender-stereotypicality are given.

Pepper:
Firefighter (male)



Myon:
Gold Digger (male)



Myon:
Florist (female)

A Experiment 3 (Adverbs – Robots) – Questionnaire: The Measured Constructs

The constructs listed in the following were measured in Experiment 3 (see Section 12.8) in addition to the assessment of participants' ambivalent sexism, normative gender role orientation, motivation to control for sexist responses, and social desirability (see Section 4.9; see Table A32 – Table A36). All items that were used in the questionnaire are listed.

Table A32

Items on Robot Anxiety (Bernotat et al.,2017; 2021)

No.	Statement: Ich habe Bedenken, dass Roboter eines Tages… [I fear, one day, robots could…]
1	… Menschen ersetzen. [… replace humans.]
2	… die Macht übernehmen. [… might take over.]
3	… einen eigenen Willen entwickeln. [… develop an own will.]
4	… Menschen kontrollieren können. [… control humans.]
5	… Menschen Schaden zufügen. [… harm humans.]
6	… Menschen angreifen. [… attack humans.]
7	… Tätigkeiten übernehmen, die eigentlich von Menschen ausgeführt werden. [… take over tasks that are commonly done by humans.]
8	… Menschen ausspionieren. [… spy humans.]

Note. 7-point Likert scales indicating 1 = I totally refuse 7 = I totally agree. The item's translation to English is inserted in parentheses below.

Table A33

Items on Robot Acceptance (Bernotat & Eyssel, 2018; Eyssel et al., 2011; Ezer, 2008; Reysen, 2005, adapted)

No.	Statement: Im Allgemeinen könnte ich mir vorstellen, einen Roboter… [In general, I could imagine…]
1	… zu besitzen. [… to own a robot.]
2	… um Rat zu bitten [… to ask a robot for advice.]
3	… kennen zu lernen. [… to get acquainted with a robot.]
4	… als Mitarbeiter/in einzustellen. [… to hire a robot.]
5	… zu nutzen. [… to use a robot.]
6	… als gute/n Freund/in zu haben. […to have a robot as a good friend.]
7	… zu testen. [… to test a robot.]
8	… zum/zur Mitbewohner/in zu haben. [… to have a robot as a housemate.]
9	… zu kaufen. [… to buy a robot.]
10	… bei mir zu Hause zu haben. [… to have a robot at home.]

Note. 7-point Likert scales indicating 1 = I totally refuse 7 = I totally agree. The item’s translation to English is inserted in parentheses below.

Table A34

Items on Agency and Communion (Bernotat et al.,2017; 2021; Eyssel & Hegel., 2012; Schneider-Düker & Kohler, 1988, adapted)

No.	Statement: Im Allgemeinen sind Roboter… [In general, robots are…]
1	… effizient. [… efficient.]
2	… geschickt. [… skilled.]
3	… kompetent. [… competent.]
4	… selbstbewusst. [… self-confident.]
5	… tüchtig. [… capable.]
6	… intelligent. [… intelligent.]
7	… aufrichtig. [… sincere.]
8	… freundlich. [… friendly.]
9	… gutmütig. [… sweet-tempered.]
10	… herzlich. [… cordially.]
11	… vertrauenswürdig. [… trustworthy.]
12	… wohlwollend. [… benevolent.]

Note. 7-point Likert scales indicating 1 = I totally refuse 7 = I totally agree. Agency: Items 1-6; Communion: Items 7-12. The item's translation to English is inserted in parentheses below.

Table A35

Items on Robot Machinelikeness and Humanlikeness (Haslam, 2006; Loughnan & Haslam, 2007; Schiffhauer, 2015, adapted)

No.	Statement: Im Allgemeinen… [In general, robots…]
1	… können Roboter große Datenmengen verarbeiten. [… can manage large datasets.]
2	… können Roboter Personen und Objekte identifizieren. [… can identify persons and objects.]
3	… sind Roboter fehleranfällig. [… are error-prone.]
4	… sind Roboter maschinenähnlich. [… are machinelike.]
5	… können Roboter in der Industrie eingesetzt werden. [… can be utilized in industry.]
6	… sind Roboter ausfallanfällig. [… are prone to technical disruption.]
7	… können Roboter in der Umgebung navigieren. [… can navigate in the environment.]
8	… sind Roboter technisch leistungsstark. [… are technically powerful.]
9	… können Roboter im häuslichen Umfeld unterstützen. [… can support in the home environment.]
10	… sind Roboter technisch. [… are technical.]
11	… können Roboter in der Produktion unterstützen. [… can support in the production process.]
12	… können Roboter die Koordinaten von Personen und Objekten speichern. [… can save coordinates of persons and objects.]
13	… können Roboter Algorithmen vervollständigen. [… can complete algorithms.]
14	… sind Roboter (technisch) hochwertig. [… are (technically) sophisticated.]
15	… sind Roboter über sich selbst bewusst. [… are self-aware.]
16	… sind Roboter eher männlich. [… are rather male.]
17	… können Roboter soziales Verhalten zeigen. [… can show social behavior.]

Note. 7-point Likert scales indicating 1 = I totally refuse 7 = I totally agree. Machinelikeness: Items 1 – 14; Humanlikeness: Items 15 – 28. The item's translation to English is inserted in parentheses below.

(continued)

Table A35

Items on Robot Machinelikeness and Humanlikeness (Haslam, 2006; Loughnan & Haslam, 2007; Schiffhauer, 2015, adapted) (continued)

No.	Statement: Im Allgemeinen… [In general, robots…]
18	… sind Roboter eher menschenähnlich. [… are rather humanlike.]
19	… können Roboter die Emotionen anderer nachempfinden. […can understand others' emotions.]
20	… verstehen Roboter moralische Fragen. [… understand moral questions.]
21	… sind Roboter eher weiblich. [… are rather female.]
22	… sind Roboter verantwortungsbewusst. [… are responsible.]
23	… sind Roboter motiviert. [… are motivated.]
24	… können Roboter menschliches Verhalten zeigen. [… can show human behavior.]
25	… können Roboter böse sein. [… can be evil.]
26	… sind Roboter ehrgeizig. [… are ambitious.]
27	… sind Roboter vernünftig. [… are reasonable.]
28	… sind Roboter ablenkbar. [… are distractable.]

Note. 7-point Likert scales indicating 1 = I totally refuse 7 = I totally agree. Machinelikeness: Items 1 – 14; Humanlikeness: Items 15 – 28. The item's translation to English is inserted in parentheses below.

Table A36

Items of the Intraindividual Differences in Anthropomorphization Questionnaire (IDAQ, Waytz et al.,2010)

No.	Statement: “In welchem Ausmaß…” [To what extent…]
1	… hat ein Auto einen freien Willen.] [… does a car have a free will.]
2	… hat ein durchschnittlicher Roboter Emotionen. […does a common robot have emotions.]
3	… haben Maschinen Intentionen. [… do machines have intentions]
4	… kann ein Fernseher Emotionen erleben. [… ca a tv experience emotions.]
5	… hat ein durchschnittlicher Computer ein Bewusstsein. [… does a common pc have consciousness.]

Note. 7-point Likert scales indicating 1 = I totally refuse 7 = I totally agree. The item’s translation to English is inserted in parentheses below.

Table A37

Items on Technology Commitment (Neyer et al., 2012)

No.	Statement
1	Hinsichtlich technischer Neuigkeiten bin ich sehr neugierig. [Regarding technical news, I am very curious.]
2	Ich finde schnell Gefallen an technischen Neuentwicklungen. [I quickly take a liking to new technical developments.]
3	Ich bin stets daran interessiert, die neusten technischen Geräte zu verwenden. [I am always interested in using the latest technical equipment.]
4	Wenn ich Gelegenheit dazu hätte, würde ich noch viel häufiger technische Produkte nutzen, als ich das gegenwärtig tue. [If I had the opportunity, I would use tech products much more often than I currently do.]
5	Im Umgang mit moderner Technik habe ich oft Angst zu versagen. [When dealing with modern technology, I am often afraid of failing.]
6	Für mich stellt der Umgang mit technischen Neuerungen zumeist eine Überforderung dar. [Dealing with technical innovations is usually too much of a challenge for me.]
7	Ich habe Angst, technische Neuentwicklungen eher kaputt zu machen, als dass ich sie richtig benutze. [I'm afraid of breaking new technical developments rather than using them properly.]
8	Den Umgang mit neuer Technik finde ich schwierig – ich kann das meistens einfach nicht. [Using new technology is difficult for me – I just cannot handle it.]
9	Ob ich erfolgreich in der Anwendung moderner Technik bin, hängt im Wesentlichen von mir ab. [Whether I am successful in the application of modern technology depends essentially on me.]
10	Es liegt in meiner Hand, ob mir die Nutzung technischer Neuentwicklung gelingt – mit Zufall oder Glück hat das wenig zu tun. [It's up to me whether I succeed in using new technical devices - this has little to do with chance or luck.]
11	Wenn ich im Umgang mit Technik Schwierigkeiten habe, hängt es schlussendlich allein von mir ab, dass ich sie löse. [If I have difficulties in dealing with technology, it finally depends on me to solve them.]
12	Das, was passiert, wenn ich mich mit technischen Neuentwicklungen beschäftigt, obliegt letztlich meiner Kontrolle. [What happens when I deal with new technical devices is under my control.]

Note. 7-point Likert scales indicating 1 = I totally refuse 7 = I totally agree. Technology Acceptance: Items 1-4; Competence in Technology Use: Items 5-8 (all items on this subscale had to be reversed); Control Over Technology Use: Items 9-1-12. The item’s translation to English is inserted in parentheses below.

Appendix B – Supplemental Results

Appendix B contains supplemental results, such as, internal consistencies of and participants' mean scores on the measured constructs, Pearson correlations, exploratory analyses, post-hoc analyses of statistical power by participants and by items per experiment, and meta-analyses by participants and by items across Experiment 1 to Experiment 4.

B Experiment 1 (Adverbs – Humans) – Internal Consistencies, Mean Scores, and Pearson Correlations

Table B1

Internal Consistencies (Cronbach's α), Means and Standard Deviations of the Measured Constructs per Speaker Voice

Speaker Voice	Construct	*α*	*M*	*SD*
Male	Benevolent Sexism	.88	3.36	1.26
	Hostile Sexism	.88	3.20	1.17
	NGRO	.90	2.84	0.83
	Social Desirability	.77	4.87	0.81
	Internal MCSR	.81	5.52	1.00
	External MCSR	.81	2.66	1.02
Female	Benevolent Sexism	.88	3.61	1.28
	Hostile Sexism	.91	3.28	1.29
	NGRO	.88	2.67	0.75
	Social Desirability	.75	4.78	0.76
	Internal MCSR	.84	5.42	1.04
	External MCSR	.80	2.49	0.93

Note. Male speaker voice: n = 44, female speaker voice: n =42. NGRO = Normative Gender Role Orientation, MCSR = Motivation to Control for Sexist Responses.

Table B2

Bivariate Pearson Correlations Between Participants' Log-Ratios for Stereotypically Male vs. Stereotypically Female Adverbs Referring to Human Targets and the Covariates

	1	2	3	4	5	6	7	8	9
1. Log-Ratios for Male Adverbs									
2. Log-Ratios for Female Adverbs	.02 [85]								
3. Speaker voice	.34** [86]	-.18† [85]							
4. Participant Gender	-.18† [84]	-.25* [83]	.15 [84]						
5. Benevolent Sexism	.19† [86]	.18† [85]	.10 [86]	-.23* [84]					
6. Hostile Sexism	-.07 [86]	.13 [85]	.03 [86]	-.14 [84]	.56*** [86]				
7. NGRO	.11 [86]	.17 [85]	-.11 [86]	-.34** [84]	.61*** [86]	.70*** [86]			
8. Social Desirability	-.09 [86]	.08 [85]	-.06 [86]	.26* [84]	-.02 [86]	-.15 [86]	-.08 [86]		
9. Internal MCSR	-.20† [86]	-.10 [85]	-.05 [86]	.29** [84]	-.42*** [86]	-.54*** [86]	-.68*** [86]	.12 [86]	
10. External MCSR	-.16 [86]	.17 [85]	-.08 [86]	.01 [84]	.29** [86]	.26* [86]	.19† [86]	-.02 [86]	-.05 [86]

Note. n indicated in brackets. NGRO = Normative Gender Role Orientation, MCSR = Motivation to Control for Sexist Responses. Participant Gender/Speaker Voice: 0 = male, 1 = female. Positive log-ratios = male character fixated, negative log-ratios = female character fixated.
***$p < .001$, **$p < .01$, *$p < .05$, $^{\dagger}p < .10$.

B Experiment 1 (Adverbs – Humans) – Exploratory Analyses

The following exploratory analyses served to explain the results reported in Section 5.2 in more detail. To do so, I examined what additional factors apart from adverb gender-stereotypicality, speaker voice, and participants' endorsement of the self-report measures might have affected their visual attention when listening to the adverbs, such as adverb connotation and participant gender. The effects of adverb connotation were explored by participants and by items. The effects of participant gender could only be tested in by-participants analyses. Moreover, participants' endorsement of the self-report measures played a crucial role to test Hypothesis 3a and Hypothesis 3b. Therefore, I explored whether participants' responses on benevolent and hostile sexism and normative gender role orientation were affected by their endorsement of internal and external motivation to control for sexist responses, and social desirability, by the speaker voice, and by participant gender.

B Experiment 1 – The Effects of Adverb Connotation, Adverb Gender-Stereotypicality, and Speaker Voice on Log-Ratios by Participants (F_1)

To explore whether adverb connotation had affected participants' eye movements when listening to the adverbs, a repeated measures MANOVA was performed with log-ratios by participants (F_1) as a function of adverb connotation and adverb gender-stereotypicality as within-participants factors and speaker voice as a between-participants factor.

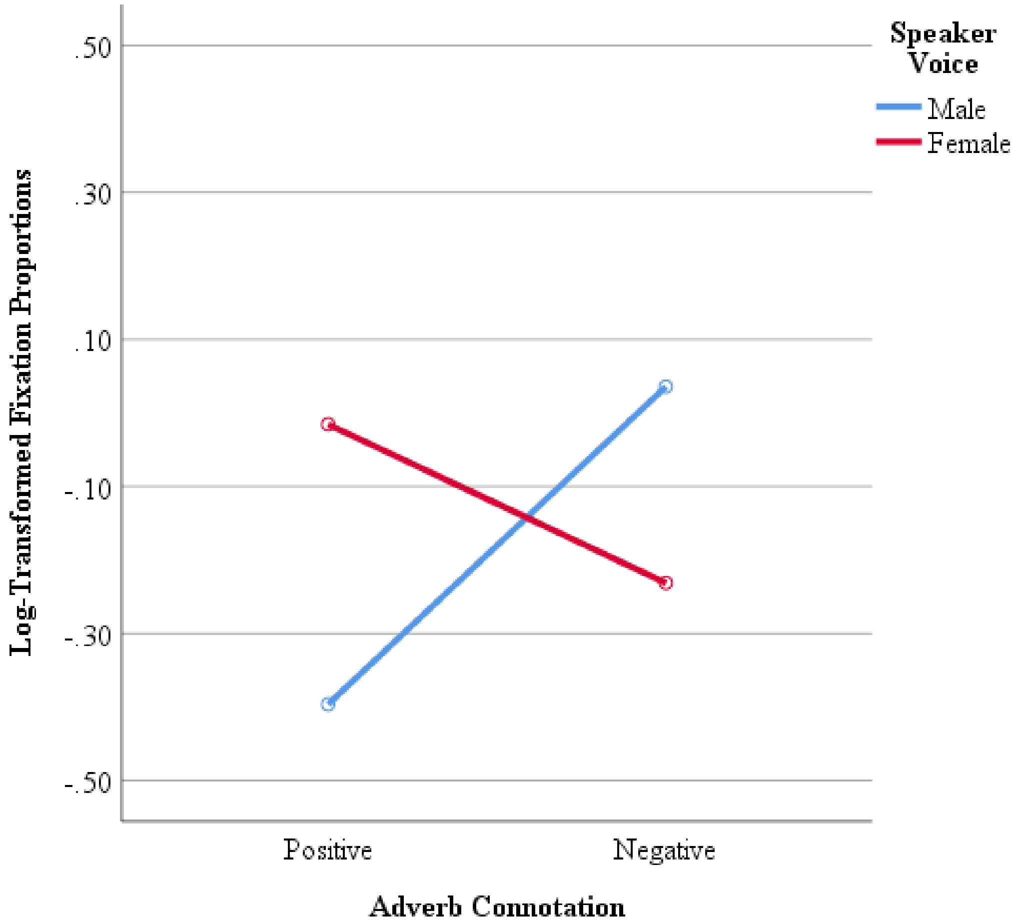
The main effects of adverb connotation, sphericity assumed: $F_1(1,84) = 0.53$, $p = .467$, $\eta_p^2 = .006$, adverb gender-stereotypicality, sphericity assumed: $F_1(1,84) = 0.73$, $p = .395$, $\eta_p^2 = .009$, and speaker voice, $F_1(1,84) = 0.20$, $p = .660$, $\eta_p^2 = .002$, were not statistically significant in by-participants analyses.

The interaction effect between adverb gender-stereotypicality and speaker voice that was statistically significant when adverb connotation was not considered, was confirmed, sphericity assumed: $F_1(1,84) = 7.39$, $p = .008$, $\eta_p^2 = .081$. Furthermore, the interaction between adverb connotation and speaker voice was statistically significant, sphericity assumed: $F_1(1,84) = 4.78$, $p = .032$, $\eta_p^2 = .054$. However, the interaction between adverb gender-stereotypicality and adverb connotation was not statistically significant, sphericity assumed: $F_1(1,84) = 0.01$, $p = .944$, $\eta_p^2 < .001$. The same accounted for the triple interaction between adverb connotation, adverb gender-stereotypicality, and speaker voice, sphericity assumed: $F_1(1,84) = 0.09$, $p = .764$, $\eta_p^2 = .001$.

To examine the interaction between adverb connotation and speaker voice in more detail, Figure B1 illustrates participants' log-ratios as a function of adverb connotation and speaker voice: As indicated by negative log-ratios, participants clearly preferred the female character over the male one when positively connoted adverbs were read by a male speaker and when negatively connoted adverbs were read by a female speaker. The preference for the female character was strongest when positively connoted adverbs were read by a male speaker. When positively connoted adverbs were uttered by a female speaker and when negatively connoted adverbs were uttered by a male speaker, log-ratios ranged around zero which implies that both characters were equally likely looked at.

Figure B1

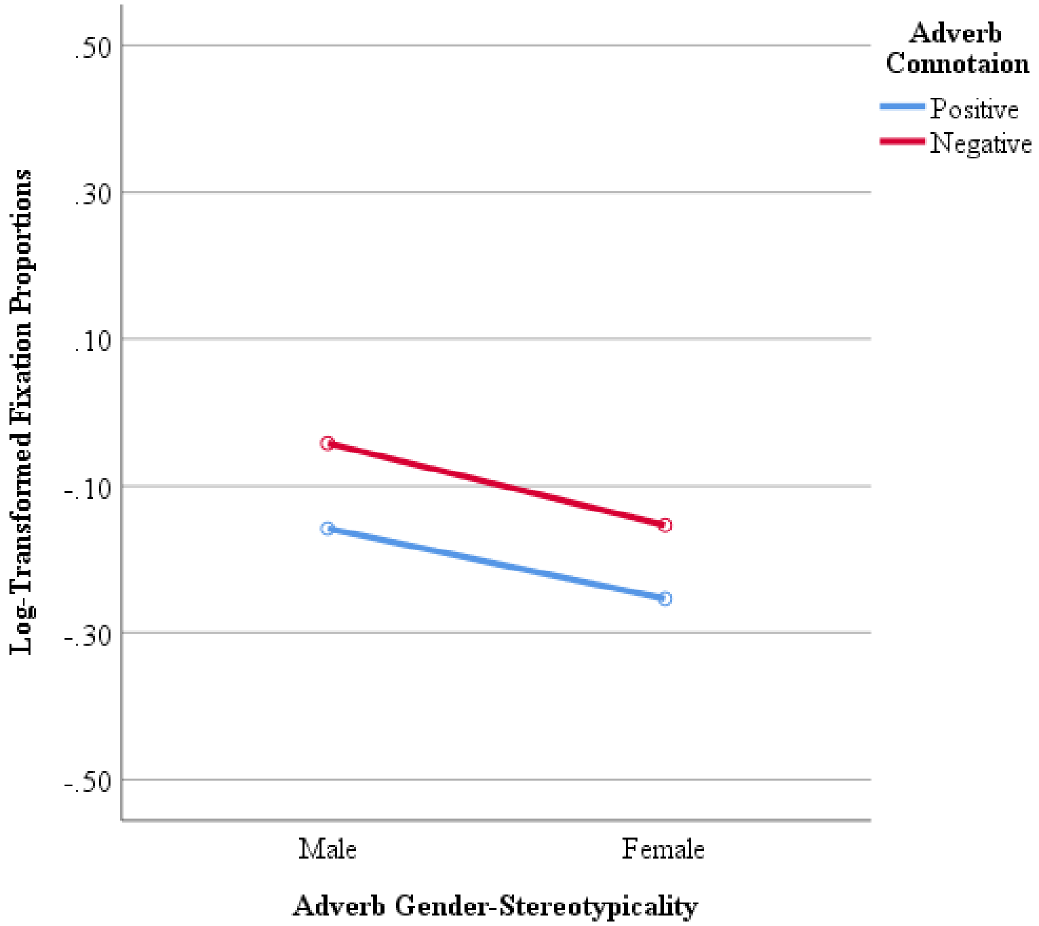
Log-Transformed Fixation Proportions as a Function of Adverb Connotation and Speaker Voice by Participants (F_1)



For reasons of completeness, the relation between adverb connotation and adverb gender-stereotypicality is also displayed in Figure B2: As indicated by negative log-ratios, the female character was more likely fixated than the male one, independent of adverb connotation and adverb gender-stereotypicality. This preference to fixate the female character was strongest for stereotypically female positively connoted adverbs and weakest for stereotypically male negatively connoted adverbs.

Figure B2

Log-Transformed Fixation Proportions as a Function of Adverb Connotation and Adverb Gender-Stereotypicality by Participants (F_1)



B Experiment 1 – The Effects of Adverb Connotation, Adverb Gender-Stereotypicality, and Speaker Voice on Log-Ratios by Items (F_2)

To explore the effect of adverb connotation by items (F_2), a repeated measures MANOVA was performed with log-ratios calculated by items as a function of adverb connotation and adverb gender-stereotypicality as between-items factors and with speaker voice as a within-items factor.

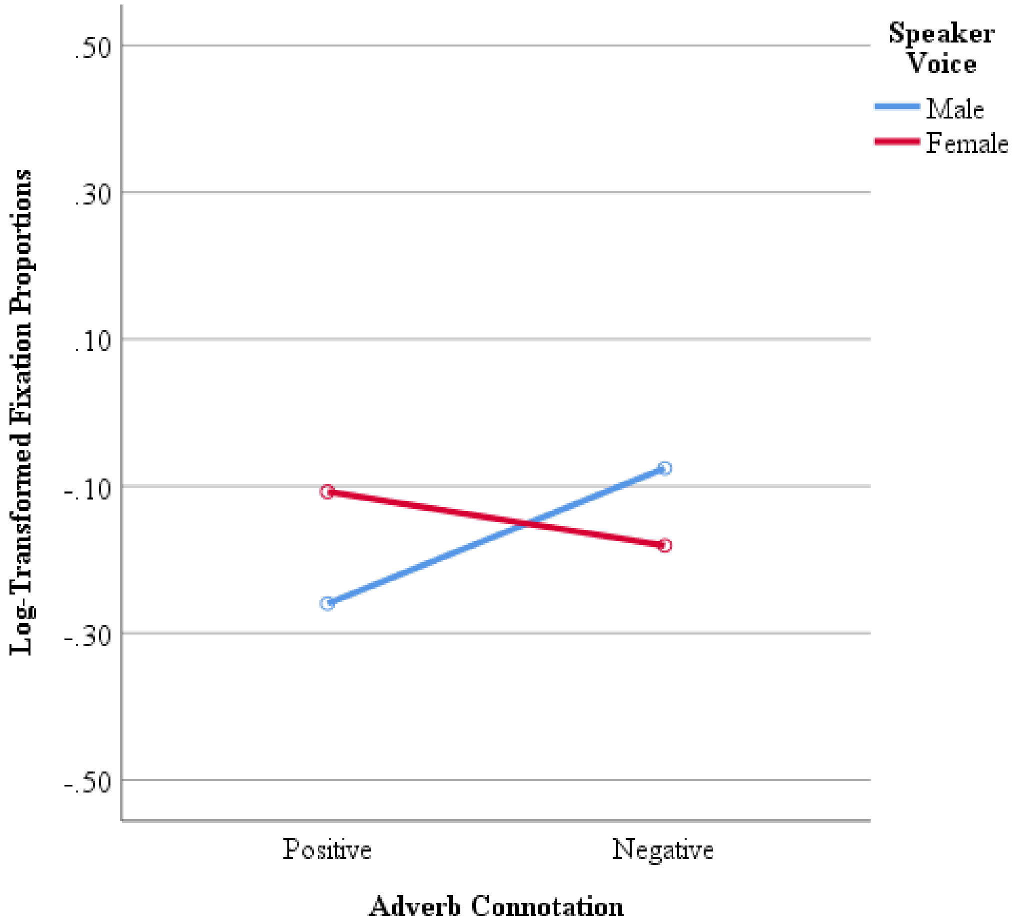
In line with by-participants results, the main effects of adverb connotation, $F_2(1,60) = 0.48$, $p = .492$, $\eta_p^2 = .008$, adverb gender-stereotypicality, $F_2(1,60) = 1.07$, $p = .305$, $\eta_p^2 = .018$, and speaker voice, sphericity assumed: $F_2(1,60) = 0.81$, $p = .777$, $\eta_p^2 = .001$, were not statistically significant.

Regarding the interaction effects, only the interaction between adverb gender-stereotypicality and speaker voice was statistically significant, sphericity assumed: $F_2(1,60) = 4.66$, $p = .035$, $\eta_p^2 = .072$. Unlike in by-participants analyses, the interaction between adverb connotation and speaker voice was not statistically significant in by-items analyses, sphericity assumed: $F_2(1,60) = 2.43$, $p = .124$, $\eta_p^2 = .039$. The same accounted for the interaction between adverb connotation and adverb gender-stereotypicality, $F_2(1,60) = 2.97$, $p = .090$, $\eta_p^2 = .047$, and for the triple interaction between adverb connotation, adverb gender-stereotypicality, and speaker voice, sphericity assumed: $F_2(1,60) = 0.16$, $p = .688$, $\eta_p^2 = .003$.

Figure B3 visualizes log-ratios calculated by items as a function of adverb connotation and speaker voice: Similar to fixation patterns by participants, the female character was most likely fixated when positively connoted adverbs were uttered by a male speaker and when negatively connoted adverbs were uttered by a female speaker. This fixation pattern was however less apparent than by-participants. Again, in line with the fixation pattern found by participants, log-ratios were negative regardless of adverb gender-stereotypicality and speaker voice which implies that the female character was generally preferred over the male one.

Figure B3

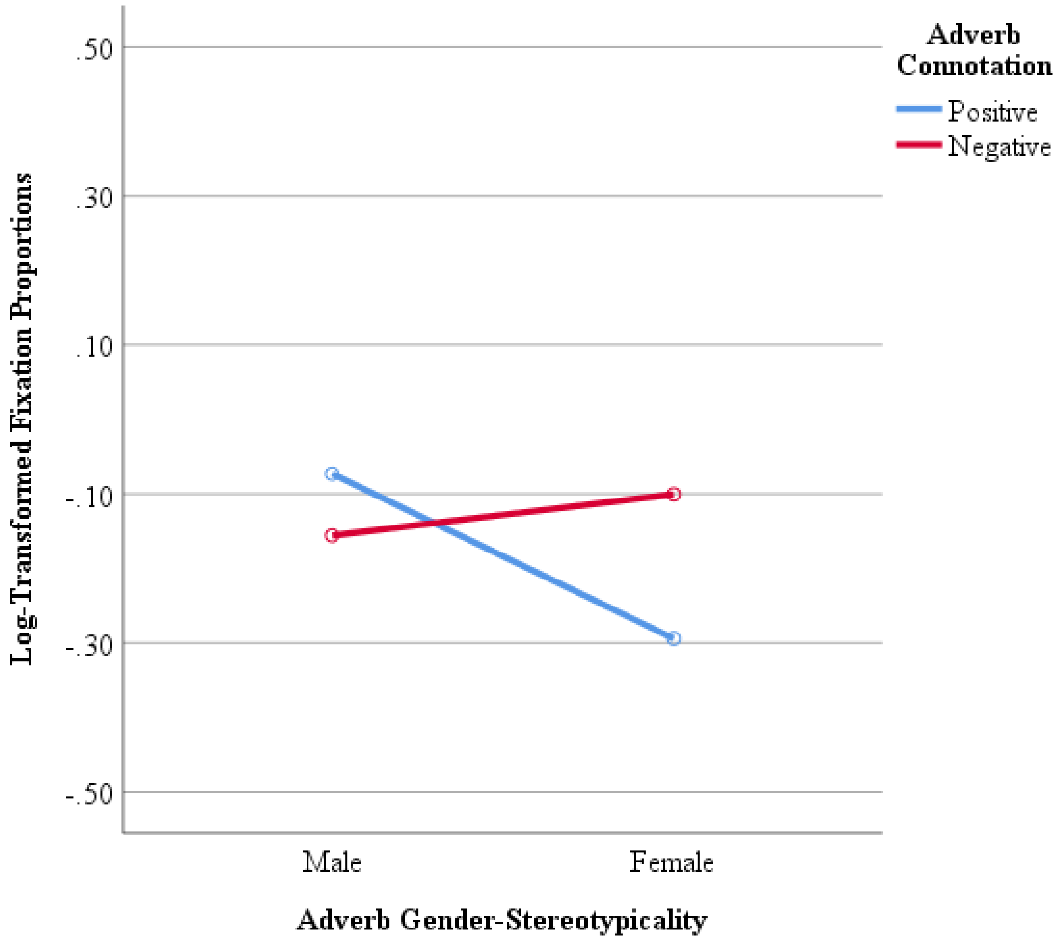
Log-Transformed Fixation Proportions as a Function of Adverb Connotation and Speaker Voice by Items (F_2)



Analogous to by-participants data, Figure B4 illustrates log-ratios calculated by items as a function of adverb connotation and adverb gender-stereotypicality: In line with fixation patterns by participants, negative log-ratios implied that the female character was more likely fixated than the male one, regardless of adverb connotation and adverb gender-stereotypicality. The preference for the female character was strongest for stereotypically female positively connoted adverbs.

Figure B4

Log-Transformed Fixation Proportions as a Function of Adverb Connotation and Adverb Gender-Stereotypicality by Items (F_2)



B Experiment 1 – The Effects of Participant Gender on Log-Ratios

Participant gender may have affected participants' visual attention when listening to gender-stereotypical adverbs and looking at male and female characters representing gender-stereotypical professions. To test this assumption, a repeated measures MANCOVA was performed with log-ratios calculated by participants as a function of adverb gender-stereotypicality as a within-participants factor and speaker voice and participant gender as between-participants factors. A balanced sample in terms of participant gender (male voice: n = 23 male, n = 20 female; female voice: n = 16 male, n = 25 female; $\chi^2(1) = 1.77$, $p = .184$) allowed to consider participant gender as a between-participants factor. This in turn enabled to explore the interactions between participant gender with each adverb gender-stereotypicality and speaker voice.

In line with analyses of Hypothesis 1 and Hypothesis 2, the main effects of adverb gender-stereotypicality, sphericity assumed: $F_1(1,79) = 1.47$, $p = .229$, $\eta_p^2 = .018$, and speaker voice, $F_1(1,79) = 2.22$, $p = .141$, $\eta_p^2 = .027$, were not statistically significant when participant gender was considered. However, the main effect of participant gender was statistically significant, $F_1(1,79) = 8.95$, $p = .004$, $\eta_p^2 = .102$. That is, not the adverbs' gender-stereotypicality and the speaker voice, but participant gender determined whether the male or the female character was looked at.

Again, in line with analyses of Hypothesis 1 and Hypothesis 2, the interaction effects between adverb gender-stereotypicality and speaker voice was statistically significant when participant gender was considered, sphericity assumed: $F_1(1,79) = 12.54$, $p = .001$, $\eta_p^2 = .137$. The interaction effects between adverb gender-stereotypicality and participant gender, sphericity assumed: $F_1(1,79) = 0.01$, $p = .918$, $\eta_p^2 < .001$, and the interaction between speaker voice and participant gender, $F_1(1,79) = 0.40$, $p = .529$, $\eta_p^2 = .005$, were not statistically significant. The same accounted for the triple interaction between adverb gender-stereotypicality, speaker voice, and participant gender, $F_1(1,79) = 3.19$, $p = .078$, $\eta_p^2 = .039$.

To explore the main effect of participant gender in more detail, male and female participants' log-ratios were displayed each as a function of adverb gender-stereotypicality and speaker voice (see Figure B5 and Figure B6).

According to Figure B5, male participants' log-ratios for stereotypically male adverbs ranged between about .50 when uttered by a female speaker and about -.30 when uttered by a male speaker. Log-ratios for stereotypically female adverbs ranged between about .10 when read by a male speaker and about -.15 when read by a female speaker. This indicates that male participants more likely looked at the character whose gender matched (vs. mismatched) the adverbs' gender-stereotypicality when listening to a female speaker voice. Reversely, when listening to a male speaker voice, male participants more likely looked at the stereotype-inconsistent character than at the gender-matching character. This fixation pattern was more apparent for stereotypically male (vs. female) adverbs.

Figure B5

Male Participants' Log-Transformed Fixation Proportions as a Function of Adverb Gender-Stereotypicality and Speaker Voice

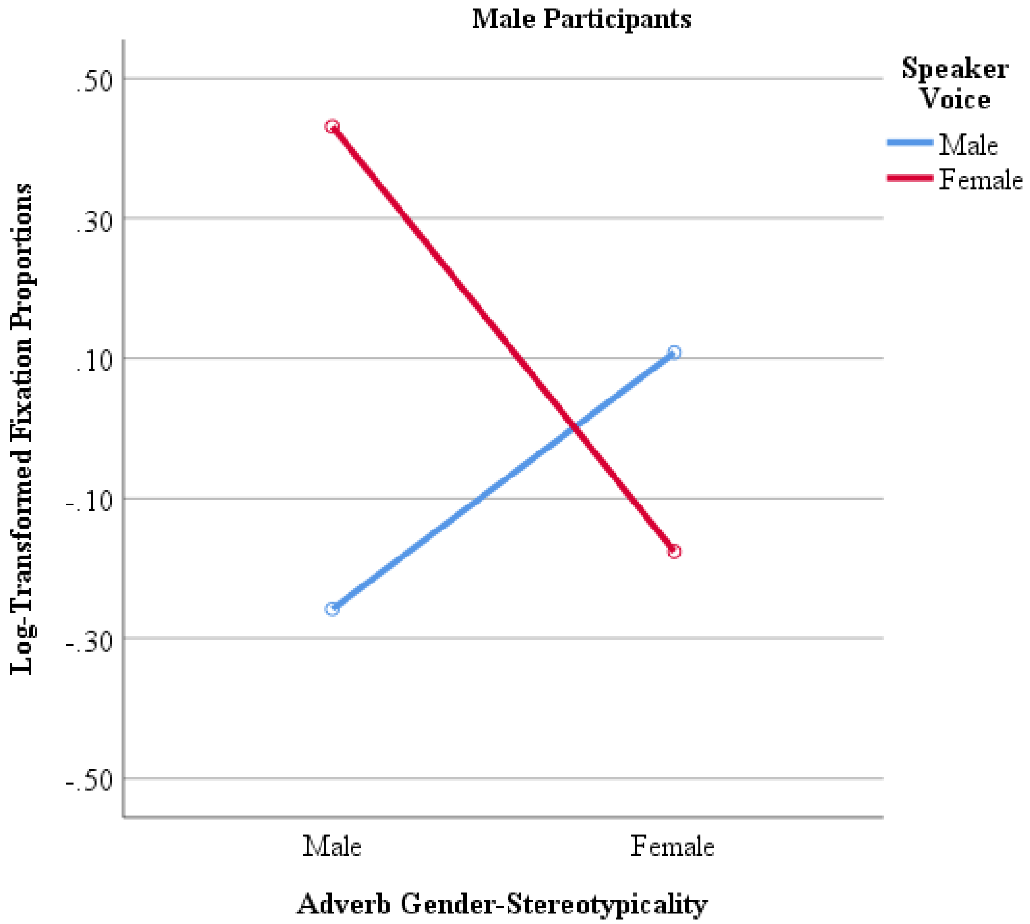
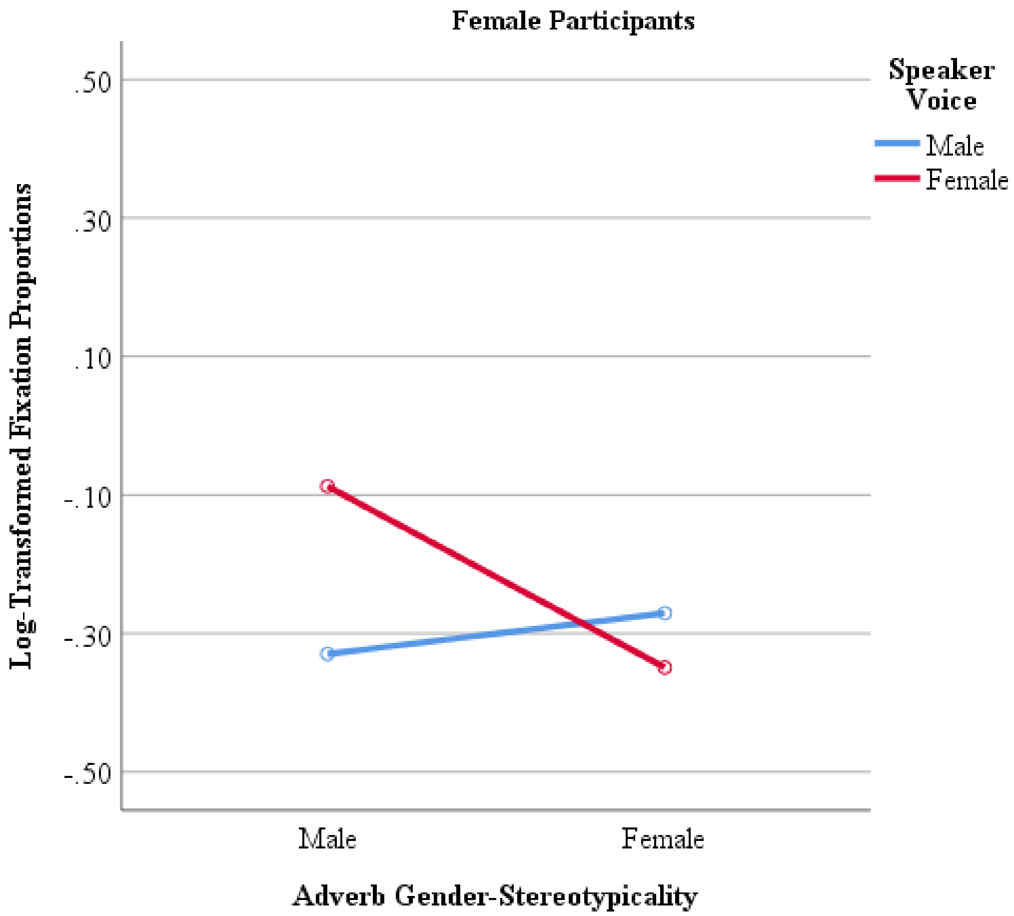


Figure B6 illustrates female participants' log-ratios as a function of adverb gender-stereotypicality and speaker voice: Similar to male participants, female participants more likely fixated the character whose gender matched (vs. mismatched) the adverbs' gender-stereotypicality when listening to a female speaker. When listening to a male speaker, they more likely fixated the stereotype-inconsistent (vs. gender-matching) character. This fixation pattern was more apparent for stereotypically male (vs. female) adverbs. Remarkably, compared to male participants, female participants' log-ratios ranged to a lesser extent and reached negative values, irrespective of adverb gender-stereotypicality and speaker voice. This indicates that female participants generally preferred the female character over the male one, while male participants looked at both characters dependent of adverb gender-stereotypicality and speaker voice.

Figure B6

Female Participants' Log-Transformed Fixation Proportions as a Function of Adverb Gender-Stereotypicality and Speaker Voice



Because participant gender had affected participants' visual attention, it was tested whether the same results would be obtained when furthermore the effects of benevolent sexism would be considered. This was done because in the context of Hypothesis 3a (see Section 5.2.3), amongst the interaction between adverb gender-stereotypicality and speaker voice, the main effect of benevolent sexism was statistically significant.

A repeated measures MANCOVA was performed with participants' log-ratios as a function of adverb gender-stereotypicality as a within-participants factor and speaker voice and participant gender as between-participants factors. Participants' endorsement of benevolent sexism was considered as a covariate.

In line with previous analyses, the main effect of participant gender, $F_1(1,78) = 5.94$, $p = .017$, $\eta_p^2 = .071$, was statistically significant. The same accounted for the main effect of benevolent sexism, $F_1(1,78) = 4.10$, $p = .046$, $\eta_p^2 = .050$, and for the interaction effect between adverb gender-stereotypicality and speaker voice, sphericity assumed: $F_1(1,78) = 12.00$, $p = .001$, $\eta_p^2 = .133$.

The main effects of adverb gender-stereotypicality, sphericity assumed: $F_1(1,78) = 0.04$, $p = .844$, $\eta_p^2 < .001$, and speaker voice, $F_1(1,78) = 1.56$, $p = .216$, $\eta_p^2 = .020$, were not statistically significant. Likewise, the interaction effect between adverb gender-stereotypicality and benevolent sexism was not statistically significant, sphericity assumed: $F_1(1,78) = 0.05$, $p = .830$, $\eta_p^2 = .001$. The same was found for the interaction effects between adverb gender-stereotypicality and participant gender, sphericity assumed: $F_1(1,78) < 0.01$, $p = .964$, $\eta_p^2 < .001$, participant gender and speaker voice, sphericity assumed: $F_1(1,78) = 1.26$, $p = .265$, $\eta_p^2 = .016$, and for the triple interaction between adverb gender-stereotypicality, speaker voice, and participant gender, sphericity assumed: $F_1(1,78) = 3.15$, $p = .080$, $\eta_p^2 = .039$.

That is, the interaction between adverb gender-stereotypicality and speaker voice, participant gender, and participants' endorsement of benevolent sexism determined which character was looked at when participants listened to the adverbs.

B Experiment 1 – The Effects of Speaker Voice, Participant Gender, Internal and External Motivation to Control for Sexist Responses, and Social Desirability on Participants' Responses on Benevolent and Hostile Sexism and NGRO

In the following, I tested whether participants' responses on benevolent and hostile sexism and normative gender role orientation had been affected by the speaker voice, participant gender, and participants' endorsement of internal and external motivation to control for sexist responses, and social desirability. That participants' endorsement of internal and external motivation to control for sexist responses and social desirability might have affected their responses on benevolent and hostile sexism and normative gender role orientation seemed plausible. The effects of speaker voice and participant gender were considered because they had seemingly played a role when listening to the adverbs. It was thus possible that they had also affected participants' responses on the self-report measures.

An independent measures MANCOVA was performed with participants' endorsement of benevolent and hostile sexism and normative gender role orientation as a function of speaker voice and participant gender as between-participants factors (male participant gender and male speaker voice were each coded by 0, female participant gender and female speaker voice were each coded by 1). Participants' endorsement of their external and internal motivation to control for sexist responses and social desirability were considered as covariates. The results are reported in Table B3.

Table B3

Results of an Independent Measures MANCOVA Considering the Effects of Speaker Voice, Participant Gender, Internal and External Motivation to Control for Sexist Responses, and Social Desirability on Participants' Responses on Benevolent and Hostile Sexism and NGRO

Independent Measure	Dependent Measure	*F*	*p*	η_p^2
Constant	NGRO	87.42	< .001	.532
	Hostile Sexism	49.04	< .001	.389
	Benevolent Sexism	16.56	< .001	.177
Speaker Voice	NGRO	2.90	.092	.036
	Benevolent Sexism	1.23	.272	.016
	Hostile Sexism	0.10	.749	.001
Participant Gender	NGRO	2.74	.102	.034
	Benevolent Sexism	2.19	.143	.028
	Hostile Sexism	0.25	.619	.003
Internal MCSR	NGRO	61.97	< .001	.446
	Hostile Sexism	31.36	< .001	.289
	Benevolent Sexism	12.82	< .001	.143
External MCSR	Benevolent Sexism	9.68	.003	.112
	Hostile Sexism	9.30	.003	.108
	NGRO	4.78	.032	.058
Social Desirability	Hostile Sexism	1.88	.174	.024
	Benevolent Sexism	0.51	.478	.007
	NGRO	0.02	.884	< .001
Speaker Voice x Participant Gender	Benevolent Sexism	5.67	.020	.069
	NGRO	4.40	.039	.054
	Hostile Sexism	4.32	.041	.053

Note. *df*(1,77) for all main effects and the interaction effect. NGRO = Normative Gender Role Orientation, MCSR = Motivation to Control for Sexist Responses.

The main effects of participants' internal motivation to control for sexist responses on their responses on normative gender role orientation, hostile sexism, and benevolent sexism were statistically significant. Pearson correlations were performed to investigate the direction of these main effects. They revealed, the higher participants' levels of internal motivation to control for sexist responses, the lower their levels of normative gender role orientation, $r(84) = -.68$, $p < .001$, hostile sexism, $r(84) = -.54$, $p < .001$, and benevolent sexism, $r(84) = -.42$, $p < .001$ (see also Table B2 for the full set of Pearson correlations).

The main effects of participants' external motivation to control for sexist responses on their responses on benevolent sexism, hostile sexism, and normative gender role orientation were statistically significant. Pearson correlations showed, the higher participants' external motivation to control for sexist responses, the higher their levels of benevolent sexism, $r(84) = .29$, $p = .007$, hostile sexism, $r(84) = .26$, $p = .016$, and normative gender role orientation, $r(84) = .19$, $p = .082$.

The main effects of speaker voice, participant gender, and social desirability on participants' responses on benevolent and hostile sexism, and normative gender role orientation were not statistically significant. Remarkably, however, the interaction effects between speaker voice and participant gender on participants' responses on benevolent sexism, hostile sexism, and normative gender role orientation were statistically significant.

To explore the interaction effect between speaker voice and participant gender on participants' responses in more detail, participants' mean ratings of benevolent sexism, hostile sexism, and normative gender role orientation were each depicted as a function of participant gender and speaker voice.

Figure B7 illustrates participants' response patterns on benevolent sexism as a function of participant gender and speaker voice: Participants indicated higher levels of benevolent sexism after having listened to a speaker who matched their own gender than after having listened to a speaker of their opposite gender. This effect was more apparent for female participants than for male participants.

Figure B7

Participants' Means on Benevolent Sexism as a Function of Participant Gender and Speaker Voice

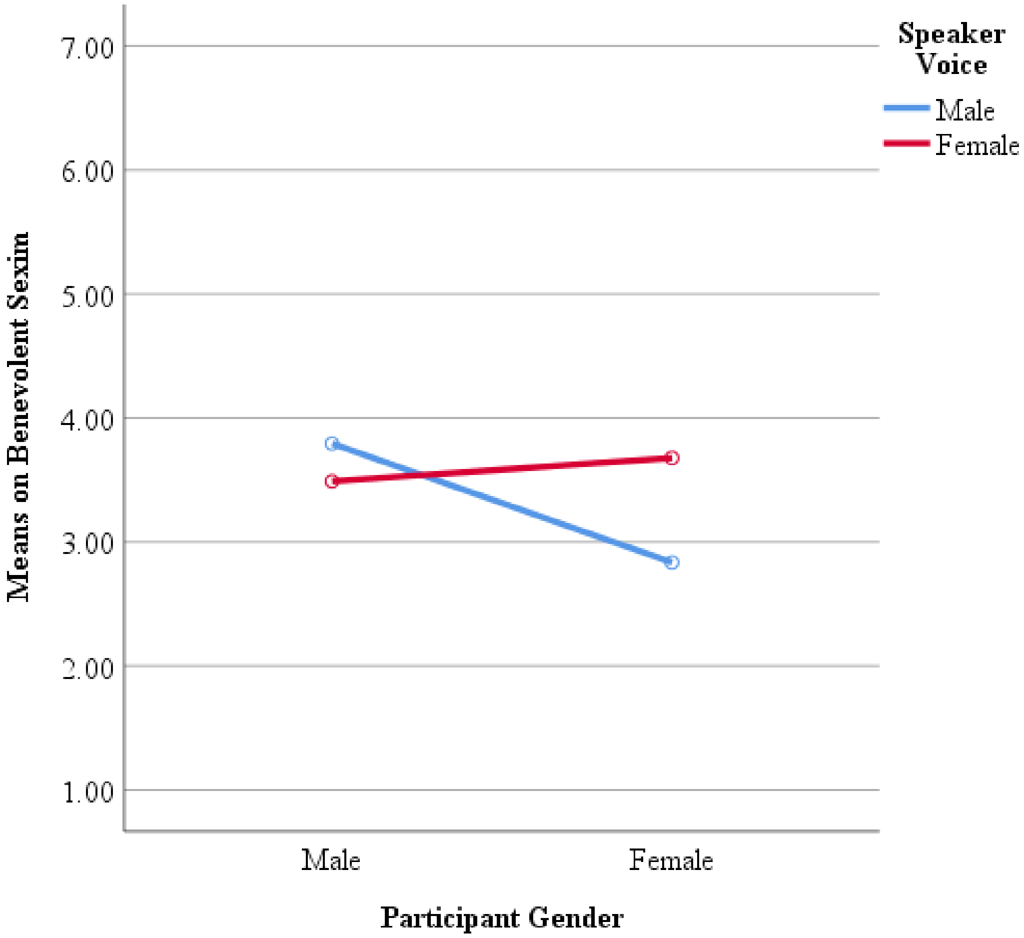


Figure B8 visualizes participants' response patterns on hostile sexism as a function of participant gender and speaker voice: Similar to response patterns for benevolent sexism, participants indicated higher levels of hostile sexism after having listened to a speaker of their own gender than after having listened to a speaker of their opposite gender. This effect was slightly more apparent for male participants than for female participants.

Figure B8

Participants' Means on Hostile Sexism as a Function of Participant Gender and Speaker Voice

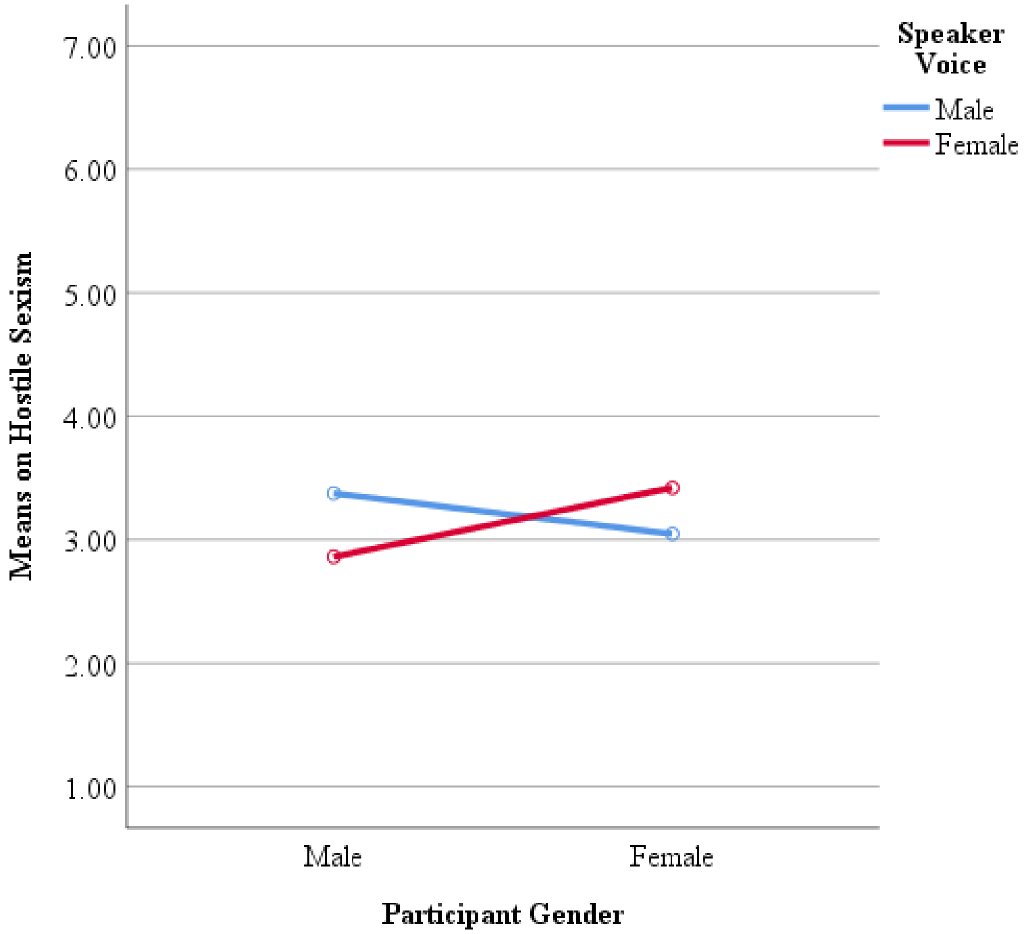
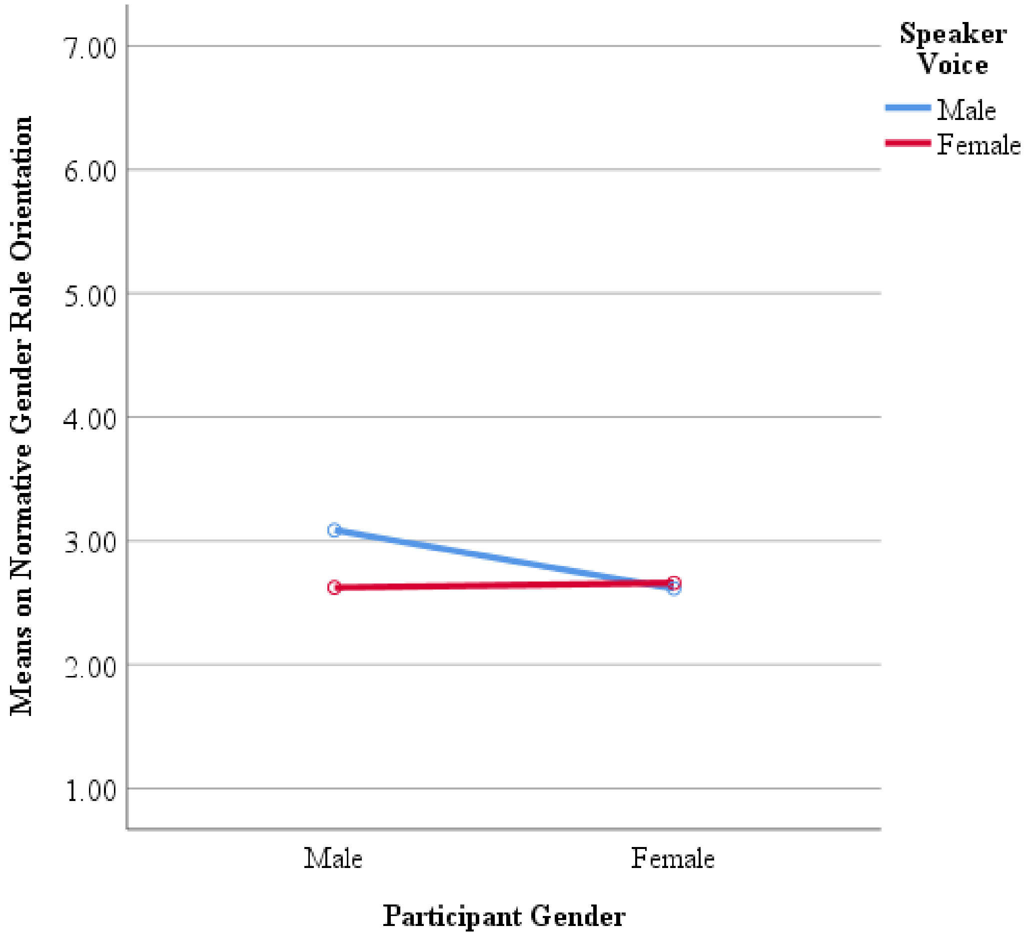


Figure B9 demonstrates participants' responses on normative gender role orientation as a function of participant gender and speaker voice: Male participants indicated slightly higher levels of normative gender role orientation after having listened to a male speaker than after having listened to a female speaker. Female participants indicated equal levels of normative gender role orientation independent of the speaker voice.

Figure B9

Participants' Means on Normative Gender Role Orientation as a Function of Participant Gender and Speaker Voice



B Experiment 2 (Main Verbs - Humans) – Internal Consistencies, Mean Scores, and Pearson Correlations

Table B4

Internal Consistencies (Cronbach's α), Means and Standard Deviations of the Measured Constructs per Speaker Voice

Speaker Voice	Construct	*α*	*M*	*SD*
Male	Benevolent Sexism	.86	3.28	1.25
	Hostile Sexism	.84	2.63	0.94
	NGRO	.85	2.38	0.63
	Social Desirability	.81	5.03	0.86
	Internal MCSR	.82	5.95	0.87
	External MCSR	.79	2.26	0.95
Female	Benevolent Sexism	.90	3.14	1.30
	Hostile Sexism	.91	2.82	1.16
	NGRO	.84	2.62	0.64
	Social Desirability	.76	4.72	0.75
	Internal MCSR	.84	5.56	0.95
	External MCSR	.84	2.37	1.00

Note. Male speaker voice: n = 37, female speaker voice: n = 41. NGRO = Normative Gender Role Orientation, MCSR = Motivation to Control for Sexist Responses.

Table B5

Bivariate Pearson Correlations Between Participants' Log-Ratios for Stereotypically Male vs. Stereotypically Female Main Verbs Referring to Human Targets and the Covariates

Construct	1	2	3	4	5	6	7	8	9
1. Log-Ratios for Male V2s									
2. Log-Ratios for Female V2s	-.33** [78]								
3. Speaker Voice	.07 [78]	-.15 [78]							
4. Participant Gender	-.04 [78]	-.14 [78]	-.03 [78]						
5. Benevolent Sexism	-.01 [78]	-.01 [78]	-.06 [78]	-.43*** [78]					
6. Hostile Sexism	-.15 [78]	-.05 [78]	.09 [78]	-.09 [78]	.49*** [78]				
7. NGRO	-.01 [78]	.12 [78]	.19† [78]	-.29** [78]	.52*** [78]	.60*** [78]			
8. Social Desirability	-.02 [78]	.02 [78]	-.20† [78]	.08 [78]	.06 [78]	-.01 [78]	-.11 [78]		
9. Internal MCSR	.02 [78]	-.03 [78]	-.21† [78]	.19† [78]	-.25* [78]	-.43*** [78]	-.64*** [78]	.20† [78]	
10. External MCSR	-.20† [78]	.08 [78]	.06 [78]	-.07 [78]	.35** [78]	.40*** [78]	.24* [78]	-.04 [78]	.01 [78]

Note. n indicated in brackets. V2s = Main Verbs. NGRO = Normative Gender Role Orientation, MCSR = Motivation to Control for Sexist Responses. Participant Gender/Speaker Voice: 0 = male, 1 = female. Positive log-ratios = male character fixated, negative log-ratios = female character fixated. ***p < .001, **p < .01, *p < .05, †p < .10.

B Experiment 2 (Main Verbs – Humans) – Exploratory Analyses

The following exploratory analyses were done to examine the results reported in Section 9.2 in more detail. To do so, I investigated whether participant gender had affected participants' visual attention when listening to the main verbs. This was done because participant gender had affected participants' visual attention when listening to gender-stereotypical adverbs in Experiment 1. It was thus possible that participant gender had also affected participants' eye gazes when listening to gender-stereotypical main verbs in Experiment 2. Moreover, participants' endorsement of the self-report measures played a crucial role to test Hypothesis 3a and Hypothesis 3b. Hostile sexism and normative gender role orientation have affected participants' visual attention when listening to the main verbs (see Section 9.2.3). Thus, analogous to Experiment 1, I explored whether participants' responses on benevolent and hostile sexism and normative gender role orientation were affected by their endorsement of internal and external motivation to control for sexist responses, social desirability, and by speaker voice and participant gender.

B Experiment 2 – The Effects of Participant Gender on Log-Ratios

To test whether participant gender had affected participants' visual attention, a repeated measures MANOVA was performed with log-ratios calculated by participants as a function of main gender-stereotypicality as a within-participants factor and speaker voice and participant gender as between-participants factors. A balanced sample in terms of participant gender (male voice: $n = 18$ male, $n = 19$ female; female voice: $n = 21$ male, $n = 20$ female; $\chi^2(1) = .05$, $p = .821$) allowed to consider participant gender as a between-participants factor. This in turn enabled to explore the interactions between participant gender with each main verb gender-stereotypicality and speaker voice.

In line with analyses on Hypothesis 1 and Hypothesis 2 in which participant gender was not considered, the main effects of main verb gender-stereotypicality, sphericity assumed: $F_1(1,74) = 0.06$, $p = .801$, $\eta_p^2 = .001$, and speaker voice, $F_1(1,74) = 0.52$, $p = .472$, $\eta_p^2 = .007$, were not statistically significant. The main effect of participant gender was not statistically significant, $F_1(1,74) = 2.10$, $p = .151$, $\eta_p^2 = .028$. That is, main verb gender-stereotypicality, speaker voice, and participant gender did not determine whether the male or the female character was looked at when listening to the main verbs.

The interaction between main verb gender-stereotypicality and speaker voice, sphericity assumed: $F_1(1,74) = 1.55$, $p = .217$, $\eta_p^2 = .021$, was not statistically significant, either.

The same accounted for the interactions between main verb gender-stereotypicality and participant gender, $F_1(1,74) = 0.35$, $p = .555$, $\eta_p^2 = .005$, and between speaker voice and participant gender, $F_1(1,74) = 0.05$, $p = .820$, $\eta_p^2 = .001$. Only the triple interaction between main verb gender-stereotypicality, speaker voice, and participant gender was statistically significant, $F_1(1,74) = 4.23$, $p = .043$, $\eta_p^2 = .054$.

To explore the triple interaction between main verb gender-stereotypicality, speaker voice, and participant gender in more detail, log-ratios for male and female participants were each displayed as a function of main verb gender-stereotypicality and speaker voice (see Figure B10 and Figure B11).

Figure B10 illustrates male participants' log-ratios as a function of main verb gender-stereotypicality and speaker voice: Male participants more likely looked at the character whose gender matched (vs. mismatched) the main verbs' gender-stereotypicality when listening to a male speaker – a speaker of their own gender. Reversely, they more likely looked at the stereotype-inconsistent (vs. gender-matching) character when the main verbs were read by a female speaker. This fixation pattern was strongest apparent for stereotypically male main verbs. When listening to stereotypically female main verbs, male participants more likely looked at the male (vs. female) character, independent of the speaker voice.

Figure B10

Male Participants' Log-Transformed Fixation Proportions as a Function of Vain Verb Gender-Stereotypicality and Speaker Voice

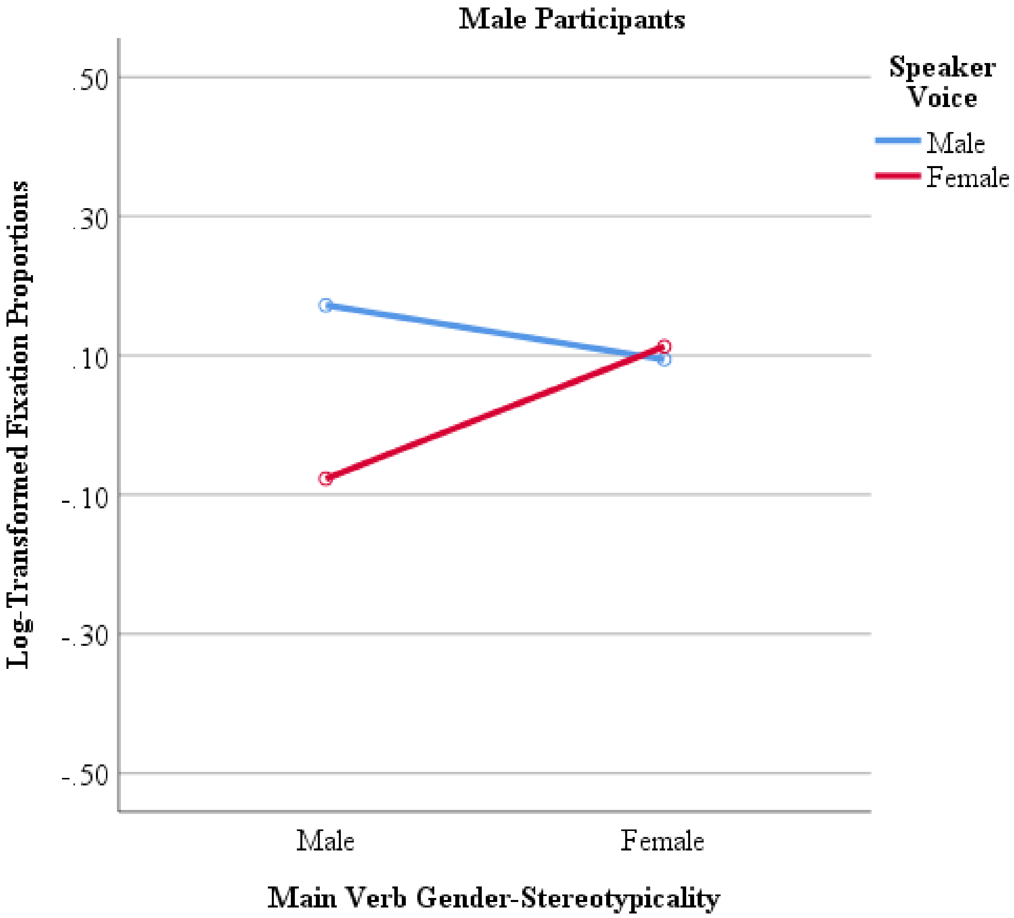
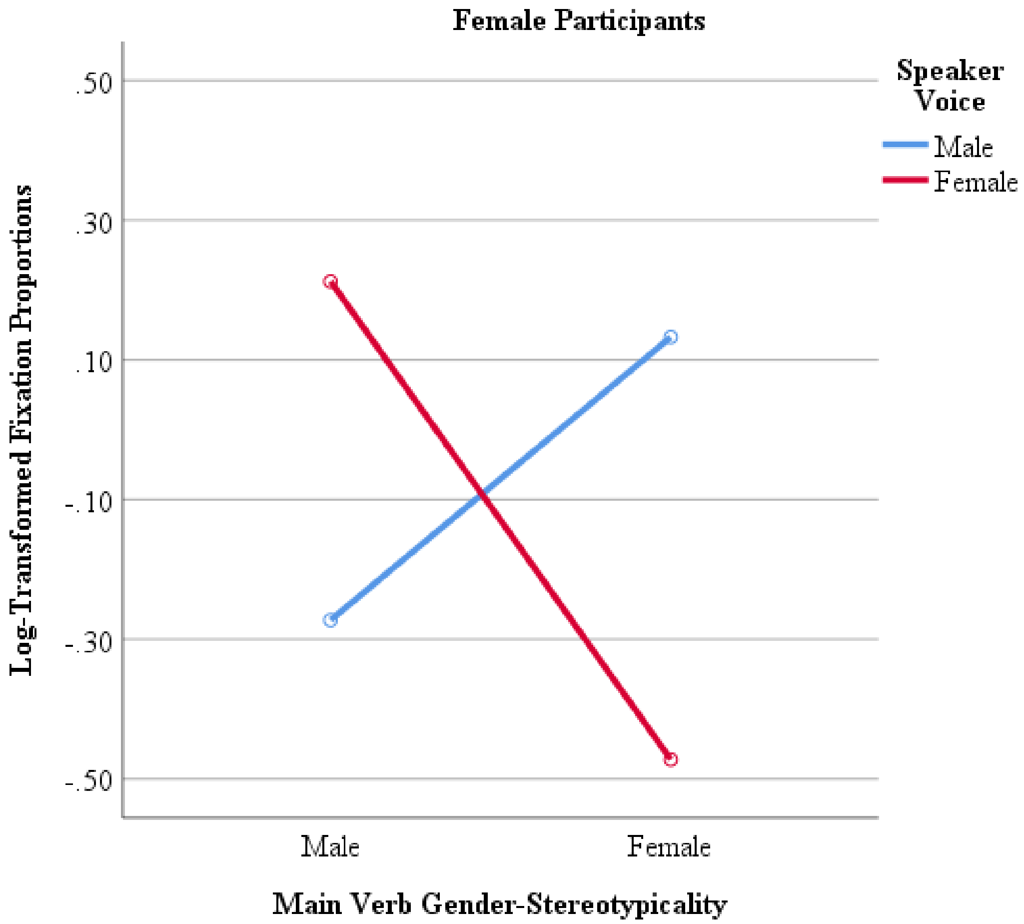


Figure B11 illustrates female participants' log-ratios as a function of main verb gender-stereotypicality and speaker voice: Similar to male participants, female participants fixated the character whose gender matched (vs. mismatched) the main verbs' gender-stereotypicality more frequently when listening to a female speaker – a speaker of their own gender. When listening to a male speaker, the character whose gender mismatched (vs. matched) the main verbs' gender-stereotypicality was more frequently fixated. Unlike male participants, this fixation pattern was apparent for stereotypically male and female main verbs.

Figure B11

Female Participants' Log-Transformed Fixation Proportions as a Function of Main Verb Gender-Stereotypicality and Speaker voice



The triple interaction between main verb gender-stereotypicality, speaker voice, and participant gender had affected participants' visual attention when listening to the main verbs. Therefore, it was tested in the following whether the same results would be obtained when furthermore the effects of hostile sexism and normative gender role orientation were considered. This was done because in the context of Hypothesis 3a, the main effects of hostile sexism and normative gender role orientation were statistically significant (see Section 9.2.3).

A repeated measures MANCOVA was performed with log-ratios calculated by participants as a function of main verb gender-stereotypicality as a within-participants factor and speaker voice and participant gender as between-participants factors. Participants' endorsement of hostile sexism and normative gender role orientation were considered as covariates. The results on Hypothesis 3a were mainly confirmed: The statistically significant main effect of hostile sexism was confirmed, $F_1(1,72) = 5.79$, $p = .019$, $\eta_p^2 = .074$. The main effect of normative gender role orientation turned out just not statistically significant, $F_1(1,72) = 3.94$, $p = .051$,

η_p^2 = .052. The same accounted for the triple interaction between main verb gender-stereotypicality, speaker voice, and participant gender, sphericity assumed: $F_1(1,72) = 3.86$, $p = .053$, η_p^2 = .051.

The main effects of main verb gender-stereotypicality, sphericity assumed: $F_1(1,72)$ = 0.41, $p = .526$, $\eta_p^2 = .006$, speaker voice, $F_1(1,72) = 0.89$, $p = .349$, $\eta_p^2 = .012$, and participant gender, $F_1(1,72) = 0.92$, $p = .342$, $\eta_p^2 = .013$, were not statistically significant.

The interaction effects between main verb gender-stereotypicality and speaker voice, sphericity assumed: $F_1(1,72) = 1.73$, $p = .193$, $\eta_p^2 = .023$, between main verb gender-stereotypicality and participant gender, sphericity assumed: $F_1(1,72) = 0.17$, $p = .684$, η_p^2 = .002, and between speaker voice and participant gender, sphericity assumed: $F_1(1,72) = 0.33$, p = .568, $\eta_p^2 = .005$, were not statistically significant. The same accounted for the interaction effects between main verb gender-stereotypicality and hostile sexism, sphericity assumed: $F_1(1,72) = 0.04$, $p = .845$, $\eta_p^2 = .001$, and main verb gender-stereotypicality and normative gender role orientation, sphericity assumed: $F_1(1,72) = 0.12$, $p = .725$, $\eta_p^2 = .002$.

That is, in an analysis in which participants' log-ratios were considered as a function of main verb gender-stereotypicality and speaker voice while furthermore the effects of participant gender and hostile sexism and normative gender role orientation were considered, only the main effect of hostile sexism determined whether the male or the female character was looked at when participants listened to the main verbs.

B Experiment 2 – The Effects of Speaker Voice, Participant Gender, Internal and External Motivation to Control for Sexist Responses, and Social Desirability on Participants' Responses on Benevolent and Hostile Sexism and NGRO

In the following, it was explored whether participants' responses on benevolent and hostile sexism and normative gender role orientation were affected by speaker voice, participant gender and participants' endorsement of internal and external motivation to control for sexist responses and social desirability.

Analogous to Experiment 1, an independent measures MANCOVA was performed with participants' endorsement of benevolent and hostile sexism and normative gender role orientation as a function of speaker voice and participant gender as between-participants factors (male participant gender and male speaker voice were each coded by 0, female participant gender and female speaker voice were each coded by 1). Participants' endorsement of external and internal motivation to control for sexist responses and social desirability were considered as covariates (see Table B6 for the results).

Table B6

Results of an Independent Measures MANCOVA Considering the Effects of Speaker Voice, Participant Gender, Internal and External Motivation to Control for Sexist Responses, and Social Desirability on Participants' Responses on Benevolent and Hostile Sexism and NGRO

Independent Measure	Dependent Measure	F	p	η_p^2
Constant	NGRO	91.04	< .001	.562
	Hostile Sexism	20.49	< .001	.224
	Benevolent Sexism	6.94	.010	.089
Speaker Voice	Benevolent Sexism	1.09	.299	.015
	NGRO	0.31	.578	.004
	Hostile Sexism	0.01	.934	0
Participant Gender	Benevolent Sexism	15.98	< .001	.184
	NGRO	3.54	.064	.048
	Hostile Sexism	0.01	.920	0
Internal MCSR	NGRO	47.61	< .001	.401
	Hostile Sexism	21.54	< .001	.233
	Benevolent Sexism	5.45	.022	.071
External MCSR	Hostile Sexism	18.90	< .001	.21
	Benevolent Sexism	13.34	< .001	.158
	NGRO	7.55	.008	.096
Social Desirability	Benevolent Sexism	2.50	.118	.034
	Hostile Sexism	1.34	.251	.019
	NGRO	0.06	.813	.001
Speaker Voice x Participant Gender	Benevolent Sexism	1.23	.272	.017
	NGRO	0.92	.341	.013
	Hostile Sexism	0.69	.408	.01

Note. *df*(1,71) for all main effects. NGRO = Normative Gender Role Orientation, MCSR = Motivation to Control for Sexist Responses.

In line with Experiment 1, the main effects of participants' endorsement of internal motivation to control for sexist responses on their levels of normative gender role orientation, hostile sexism, and benevolent sexism were statistically significant. Pearson correlation were calculated to investigate the direction of these main effects. They revealed that the higher participants' scored on internal motivation to control for sexist responses, the lower levels of normative gender role orientation, $r(76) = -.64$, $p < .001$, hostile sexism, $r(76) = -.43$, $p < .001$, and benevolent sexism, $r(76) = -.25$, $p = .028$, they reported.

The main effects of participants' endorsement of external motivation to control for sexist responses on their levels of hostile sexism, benevolent sexism, and normative gender role orientation were statistically significant. Pearson correlations showed that the higher participants' levels of external motivation to control for sexist responses, the higher levels of hostile sexism, $r(76) = .40$, $p < .001$, benevolent sexism, $r(76) = .35$, $p = .002$, and normative gender role orientation, $r(76) = .24$, $p = .032$, they indicated.

Regarding the effects of participant gender, only the main effect of participant gender on benevolent sexism was statistically significant. Pearson correlations revealed that female participants indicated lower levels of benevolent sexism than males, $r(76) = -.43$, $p < .001$.

The main effects of speaker voice and social desirability on participants' responses on benevolent and hostile sexism and normative gender role orientation were not statistically significant. Unlike Experiment 1, the interaction effects between participant gender and speaker voice on participants' levels of benevolent and hostile sexism and normative gender role orientation were not statistically significant (see Table B6 for the results; see Table B5 for the full set of Pearson correlations).

B Experiment 3 (Adverbs – Robots) – Internal Consistencies, Mean Scores, and Pearson Correlations

Table B7

Internal Consistencies (Cronbach's α), Means and Standard Deviations of the Measured Constructs per Speaker Voice

Speaker Voice	Construct	*α*	*M*	*SD*
Male	Benevolent Sexism	.87	3.02	1.23
	Hostile Sexism	.89	2.45	0.98
	NGRO	.78	2.26	0.54
	Social Desirability	.65	4.84	0.58
	Internal MCSR	.82	5.82	0.80
	External MCSR	.75	2.38	0.82
	Robot Anxiety	.89	4.09	1.50
	Robot Acceptance	.89	4.48	1.39
	Agency	.67	4.82	1.05
	Communion	.79	3.04	1.18
	Robot Machinelikeness	.88	6.13	0.65
	Robot Humanlikeness	.87	2.70	1.06
	TC – Acceptance	.85	4.09	1.37
	TC – Competence	.88	5.89	1.15
	TC – Control	.92	4.70	1.58
	IDAQ	.49	1.59	0.58
Female	Benevolent Sexism	.94	3.17	1.58
	Hostile Sexism	.92	2.72	1.20
	NGRO	.88	2.56	0.78
	Social Desirability	.86	4.66	0.96
	Internal MCSR	.84	5.50	0.99
	External MCSR	.71	2.72	0.86
	Robot Anxiety	.87	4.12	1.41
	Robot Acceptance	.87	3.98	1.37
	Agency	.56	4.68	0.91
	Communion	.73	3.05	1.04
	Robot Machinelikeness	.67	6.08	0.47
	Robot Humanlikeness	.80	2.48	0.83
	TC – Acceptance	.92	3.60	1.57
	TC – Competence	.91	5.45	1.49
	TC – Control	.66	5.25	0.89
	IDAQ	.72	1.58	0.74

Note. Male speaker voice: $n = 38$, female speaker voice: $n = 34$. NGRO = Normative Gender Role Orientation, MCSR = Motivation to Control for Sexist Responses, TC = Technology Commitment, IDAQ = Individual Differences in Anthropomorphism Questionnaire.

Table B8

Bivariate Pearson Correlations Between Participants' Log- Ratios for Stereotypically Male vs. Stereotypically Female Adverbs Referring to Robot Targets and the Covariates

Construct	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. Log-Ratios for Male Adverbs																			
2. Log-Ratios for Female Adverbs	-.10 [72]																		
3. Speaker Voice	-.16 [72]	-.13 [72]																	
4. Participant Gender	.14 [72]	-.15 [72]	-.03 [72]																
5. Benevolent Sexism	-.17 [72]	.04 [72]	.05 [72]	-.18 [72]															
6. Hostile Sexism	-.02 [72]	-.03 [72]	.13 [72]	.00 [72]	.54*** [72]														
7. NGRO	-.13 [72]	.03 [72]	.22† [72]	-.17 [72]	.54*** [72]	.67*** [72]													
8. Social Desirability	-.13 [72]	.17 [72]	-.12 [72]	.03 [72]	.18 [72]	-.07 [72]	-.09 [72]												
9. Internal MCSR	.09 [72]	-.01 [72]	-.18 [72]	.16 [72]	-.45*** [72]	-.61*** [72]	-.62*** [72]	.11 [72]											
10. External MCSR	-.19 [72]	.28* [72]	.20† [72]	-.15 [72]	.31** [72]	.35** [72]	.49*** [72]	-.07 [72]	-.28* [72]										

Note. n indicated in brackets. NGRO = Normative Gender Role Orientation, MCSR = Motivation to Control for Sexist Responses, TC = Technology Commitment, IDAQ = Individual Differences in Anthropomorphism Questionnaire. Participant Gender/Speaker Voice: 0 = male, 1 = female. Positive log-ratios = male character fixated, negative log-ratios = female character fixated. $^{***}p < .001$, $^{**}p < .01$, $^{*}p < .05$, $^{\dagger}p < .10$.

(continued)

Table B8

Bivariate Pearson Correlations Between Participants' Log- Ratios for Stereotypically Male vs. Stereotypically Female Adverbs Referring to Robot Targets and the Covariates (continued)

Construct	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
11. Robot Anxiety	.04	-.11	.01	-.05	.10	.06	.05	-.22†	.01	.10									
	[72]	[72]	[72]	[72]	[72]	[72]	[72]	[72]	[72]	[72]									
12. Robot Acceptance	-.23†	.03	-.18	-.28*	-.01	.04	-.01	-.30*	-.20	< .01	-.08								
	[72]	[72]	[72]	[72]	[72]	[72]	[72]	[72]	[72]	[72]	[72]								
13. Agency	-.13	.02	-.07	-.06	.19	.05	.07	.16	.03	.10	-.10	.20†							
	[72]	[72]	[72]	[72]	[72]	[72]	[72]	[72]	[72]	[72]	[72]	[72]							
14. Communion	-.18	-.10	.01	-.03	.09	.05	-.01	.02	.05	.21†	-.02	.21†	.55***						
	[72]	[72]	[72]	[72]	[72]	[72]	[72]	[72]	[72]	[72]	[72]	[72]	[72]						
15. Robot Machinelikeness	-.06	-.03	-.04	< .01	-.02	-.14	-.30*	-.16	.13	-.20†	.04	.18	.40***	.02					
	[72]	[72]	[72]	[72]	[72]	[72]	[72]	[72]	[72]	[72]	[72]	[72]	[72]	[72]					
16. Robot Humanlikeness	-.23†	.09	-.12	-.04	.24*	.18	.07	.05	< .01	.26*	.03	.21†	.49***	.73***	.10				
	[72]	[72]	[72]	[72]	[72]	[72]	[72]	[72]	[72]	[72]	[72]	[72]	[72]	[72]	[72]				
17. TC - Acceptance	-.16	.12	-.17	-.36**	.25*	.22†	.24*	-.16	-.23†	.36**	-.11	.45***	.15	.23†	-.05	.27*			
	[71]	[71]	[71]	[71]	[71]	[71]	[71]	[71]	[71]	[71]	[71]	[71]	[71]	[71]	[71]	[71]			
18. TC - Agency	-.16	.20†	-.17	-.31**	.16	-.04	-.08	.02	-.04	-.11	-.10	.27*	.05	-.09	.16	.03	.32**		
	[71]	[71]	[71]	[71]	[71]	[71]	[71]	[71]	[71]	[71]	[71]	[71]	[71]	[71]	[71]	[71]	[71]		
19. TC - Control	.15	.04	.21†	-.23†	.10	.13	.03	.04	-.01	.15	.11	-.08	-.04	-.26*	.04	-.20†	.20†	.14	
	[71]	[71]	[71]	[71]	[71]	[71]	[71]	[71]	[71]	[71]	[71]	[71]	[71]	[71]	[71]	[71]	[71]	[71]	
20. IDAQ	-.02	.11	.00	-.04	.18	.08	-.01	.14	-.06	.27*	.10	.14	.24*	.46***	-.02	.59***	.24*	.26*	-.06
	[71]	[71]	[71]	[71]	[71]	[71]	[71]	[71]	[71]	[71]	[71]	[71]	[71]	[71]	[71]	[71]	[71]	[71]	[71]

Note. n indicated in brackets. NGRO = Normative Gender Role Orientation, MCSR = Motivation to Control for Sexist Responses, TC = Technology Commitment, IDAQ = Individual Differences in Anthropomorphism Questionnaire. Participant Gender/Speaker Voice: 0 = male, 1 = female. Positive log-ratios = male character fixated, negative log-ratios = female character fixated. ****p* < .001, ***p* < .01, **p* < .05, †*p* < .10.

B Experiment 3 (Adverbs – Robots) – Exploratory Analyses

Exploratory analyses were conducted to explain the results in more detail. To do so, participants' ability to recognize the robots' alleged gender and professions and their ability to imagine robots performing the presented tasks and professions were explored. The purpose of doing these analyses was two-fold: First, participants' perceptions of the robot stimuli likely determined whether participants could link the sentences' content to the visual robot stimuli. They might therefore have affected participants' eye gazes within the adverb and the NP2 region. Second, these analyses served to gain more detailed information about participants' views of robots in general. Therefore, participants' preferences for robot use, robot acceptance, robot anxiety, and technology commitment were furthermore enquired (see Table B7).

B Experiment 3 – Participants' Ability to Recognize the Robots' Professions and Gender and to Imagine Robots Performing the Presented Tasks and Professions

Participants' ability to recognize the robots' alleged professions and gender was deemed important to link the sentences' content to the visual robot stimuli. The same accounted for participants' ability to imagine robots performing the presented tasks and professions. Therefore, participants' evaluations of the robots' recognizability in terms of profession and male and female robot gender were assessed. Moreover, participants' ability to imagine robots performing the presented tasks and professions were measured using 7-point Likert scales. To quantify participants' evaluations, one-sample *t*-tests were performed against the scale midpoint of 4 (see Table B9)[34]:

Participants reported to have recognized the robots' professions. Furthermore, they judged the robots as portraying the presented professions in a typical manner and stated to have perceived one of the robots within a pair as male and the other as female. These findings suggest that participants could identify the robot targets which is also in line with their fixation patterns within the NP2 region (see Section 13.4). Participants' ability to imagine robots performing the presented tasks was moderate, while their ability to imagine robots performing the presented professions was low relative to the scale midpoint.

Table B9

Results of One-Sample t-Tests Against the Scale Midpoint of 4 on Participants' Evaluations of the Robots' Recognizability in Terms of Profession and Gender, the Professions' Typicality, and Participants' Ability to Imagine Robots to Perform the Presented Tasks and Professions

Item	*M*	*SD*	*t*	*df*	*p*	Cohen's *d*
Recognizability of the Robots' Professions	4.72	1.56	3.93	71	< .001	0.46
Typicality of the Professions' Portraits	5.14	1.37	7.07	71	< .001	0.83
Male Robot Gender	5.92	1.11	14.60	70	< .001	1.73
Female Robot Gender	5.72	1.32	10.95	70	< .001	1.30
Imaginability of Tasks	3.70	1.69	-1.47	70	.146	-0.17
Imaginability of Professions	3.45	1.75	-2.65	70	.010	-0.31

Note. High means indicate high agreement to the respective construct.

B Experiment 3 – Tasks and Professions for Which Robots Would be Used

Complementary to participants' ability to imagine robots to perform the presented tasks and professions, I was interested in knowing for what kind of tasks and professions they would use robots. Therefore, using open-response formats, participants were asked to name a task and a profession for which they would mainly use robots. Their responses are reported in the following:

***Tasks* Participants Indicated for Which They Would Prefer to Use Robots:** Constructing machines, cashing, elderly care, nursing, designing things (after humans' instruction), protecting people, steering machines in industry (especially in dangerous settings), book keeping, sorting things, stocking shelves (e.g., in a super marked), running errands, mounting things, data management, data storage, domestic work (e.g., cleaning, garbage disposal, shopping), transporting heavy things, high-precision work, and traffic monitoring.

***Professions* Participants Indicated for Which They Would Prefer to Use Robots:** Industrial worker, secretary work, elderly care, nurse, construction worker, construction manager, shop assist, bank clerk, butler, truck driver, cleaner, florist, craftsman, domestic assistant, cleaner, storekeeper, painter, pilot, train driver, taxi driver, mechatronics engineer, steel worker, carpenter, welder, soldier, explosives expert, medical assistant, and crossing guard (see also Bernotat & Eyssel, 2018; Bernotat et al.,2021 for similar findings).

To quantify participants' preferences for robot use, participants had to judge the task and the profession they had named as *safe* vs. *dangerous*, *interesting* vs. *boring*, *female* vs. *male*, *socially interactive* vs. *socially isolated*, *demanding* vs. *simple*. One-sample *t*-tests against the scale midpoint of 4 were performed (see Table B10)[34]. Robots were deemed suitable for rather safe[42], boring, stereotypically male, and socially isolated *tasks* and for rather safe *professions*.

Table B10

Results of One-Sample t-Tests Against the Scale Midpoint of 4 on Participants' Evaluations of the Tasks and Professions for Which They Would Use Robots

Category	Item	*M*	*SD*	*t*	*df*	*p*	Cohen's *d*
Tasks	Safe vs. Dangerous	2.51	1.70	-7.44	71	< .001	0.88
	Interesting vs. Boring	4.88	1.93	3.85	71	< .001	0.45
	Stereot. Female vs. Male	4.21	0.67	2.64	70	.010	0.31
	Interactive vs. Isolated	5.11	1.95	4.84	71	< .001	0.57
	Demanding vs. Simple	4.21	1.94	0.91	71	.364	0.11
Professions	Safe vs. Dangerous	3.27	1.79	-3.45	70	.001	0.41
	Interesting vs. Boring	4.20	1.97	0.84	70	.402	0.10
	Stereot. Female vs. Male	4.06	0.83	0.57	69	.567	0.07
	Interactive vs. Isolated	4.24	2.10	0.96	70	.340	0.11
	Demanding vs. Simple	3.68	1.97	-1.39	70	.170	0.16

Note. 7-point semantic differentials were used (e.g., 1 = "safe", 7 = "dangerous"). High means thus indicate high agreement to the latter mentioned attribute.

[42] Because participants indicated to use robots in dangerous settings, perhaps they considered 'safe' as safe for themselves, so that robots could not harm them or prevent them from getting into a dangerous situation.

B Experiment 3 – The Effects of Adverb Connotation, Adverb Gender-Stereotypicality, and Speaker Voice on Log-Ratios by Participants (F_1)

Similar to Experiment 1, it was tested whether adverb connotation had affected participants' visual attention when robots were displayed in Experiment 3. To do so, a repeated measures MANOVA was performed analogous to Experiment 1. Log-ratios calculated by participants (F_1) were considered as a function of adverb connotation and adverb gender-stereotypicality as within-participants factors and speaker voice as a between-participants factor.

The main effects of adverb connotation, sphericity assumed: $F_1(1,70) = 0.20$, $p = .655$, $\eta_p^2 = .003$, adverb gender-stereotypicality, sphericity assumed: $F_1(1,70) = 2.49$, $p = .119$, $\eta_p^2 = .034$, and speaker voice, $F_1(1,70) = 2.48$, $p = .120$, $\eta_p^2 = .034$, were not statistically significant in analyses conducted by participants. Likewise, the interactions between adverb connotation and adverb gender-stereotypicality, sphericity assumed: $F_1(1,70) = 0.27$, $p = .602$, $\eta_p^2 = .004$, adverb connotation and speaker voice, $F_1(1,70) = 1.96$, $p = .166$, $\eta_p^2 = .027$, and adverb gender-stereotypicality and speaker voice, sphericity assumed: $F_1(1,70) = 0.67$, $p = .416$, $\eta_p^2 = .009$, were not statistically significant. The same accounted for the triple interaction between adverb connotation, adverb gender-stereotypicality, and speaker voice, sphericity assumed: $F_1(1,70) = 0.61$, $p = .437$, $\eta_p^2 = .009$.

That is, in by-participant analyses, adverb connotation, amongst adverb gender-stereotypicality and speaker voice, had no considerable effects on participants' visual attention when listening to the adverbs. To descriptively explore participants' fixation patterns as a function of adverb connotation, adverb gender-stereotypicality, and speaker voice, log-ratios were displayed per male and female speaker voice (see Figure B12 and Figure B13).

Figure B12 visualizes participants' log-ratios when listening to a male speaker voice as a function of adverb gender-stereotypicality and adverb connotation: At a descriptive level, participants tended to look at the robot whose gender matched (vs. mismatched) the adverbs' gender-stereotypicality. However, log-ratios were positive, except for stereotypically female negatively connoted adverbs. This indicates that, when listening to a male speaker voice, the male robot was more likely fixated than the female one, except when stereotypically female negatively connoted adverbs were encountered. In this case, the female robot was more likely looked at than the male one. When listening to a male speaker voice, the preference for the male robot was most apparent when stereotypically male positively connoted adverbs were uttered.

Figure B12

Log-Transformed Fixation Proportions in the Male Speaker Voice Condition as a Function of Adverb Connotation and Adverb Gender-Stereotypicality by Participants (F_1)

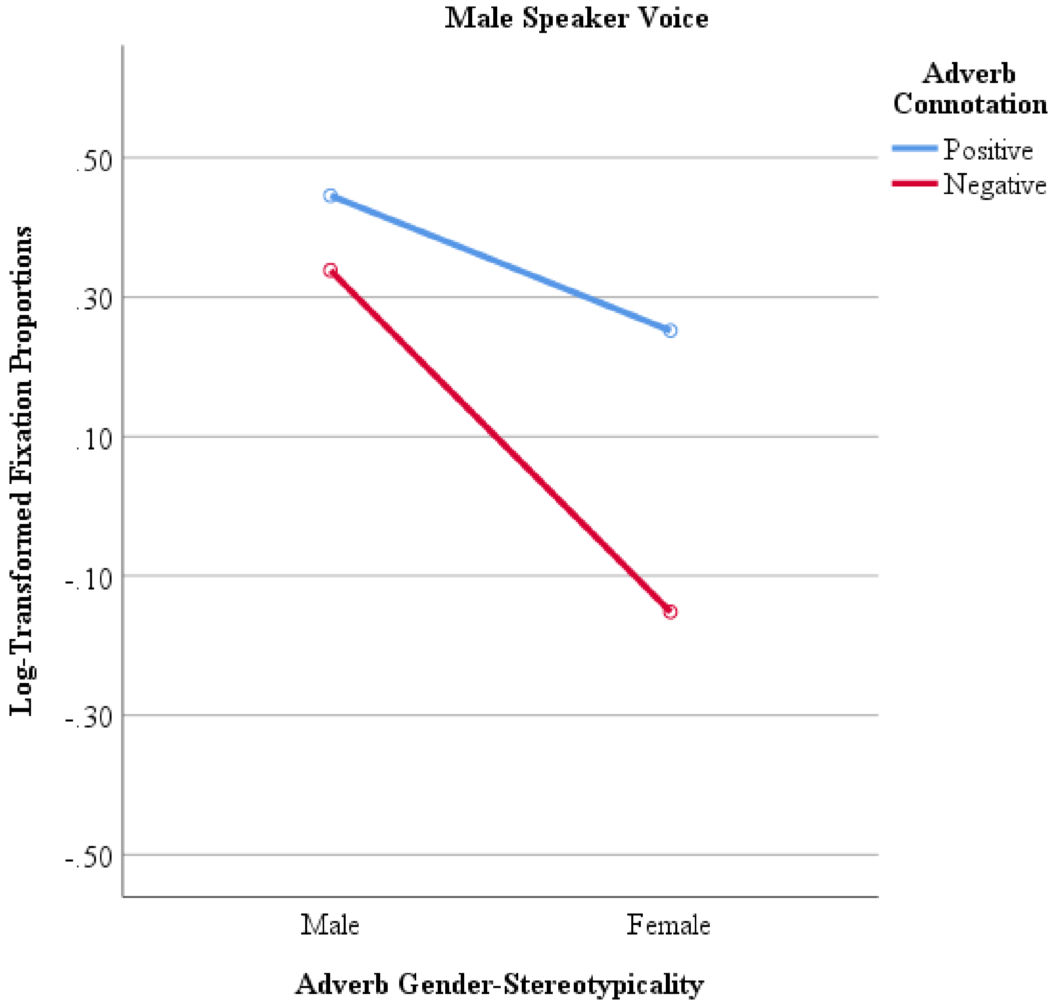
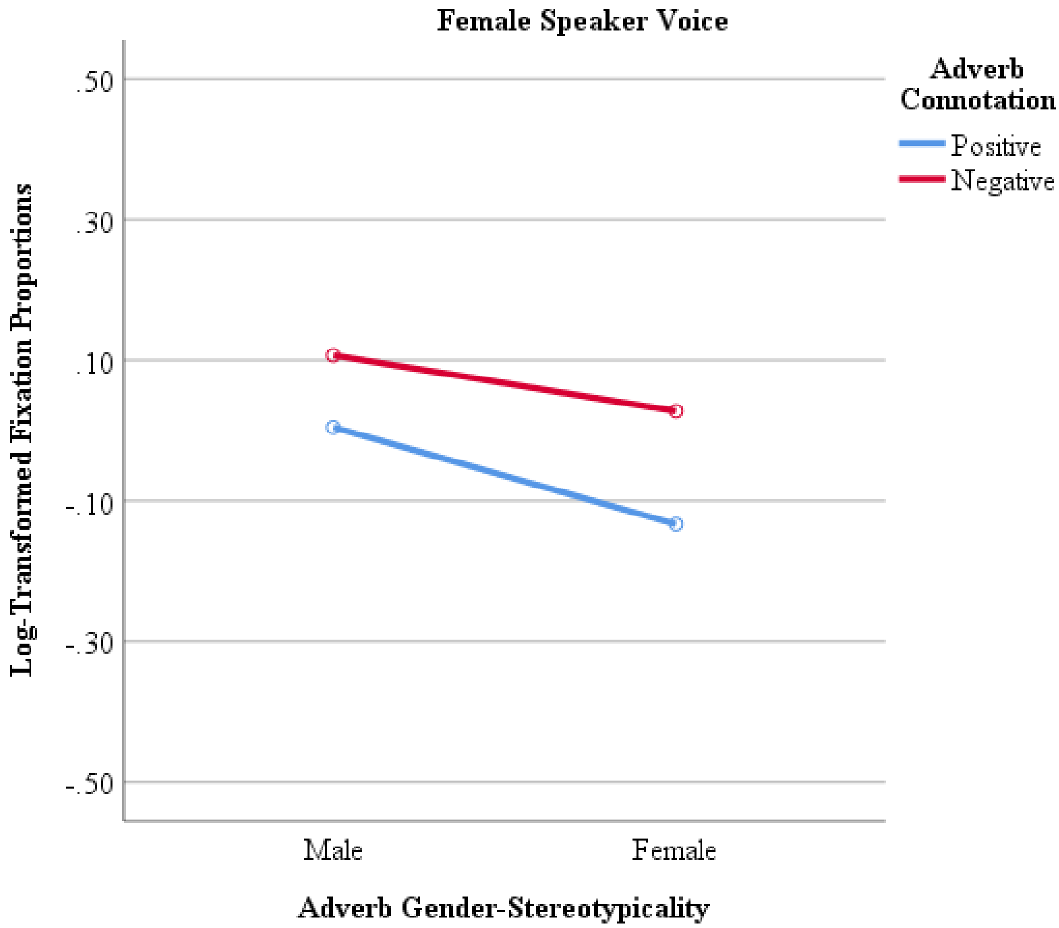


Figure B13 illustrates participants' log-ratios when listening to a female speaker voice as a function of adverb gender-stereotypicality and adverb connotation: In line with the fixation pattern when listening to a male speaker voice, at a descriptive level, participants tended to look at the robot whose gender matched (vs. mismatched) the adverbs' gender-stereotypicality. Log-ratios when listening to a female speaker reached negative values only when stereotypically female positively connoted adverbs were uttered. This indicates that, when listening to a female speaker voice, there was a tendency to look at the female robot only when stereotypically female positively connoted adverbs were uttered. Compared to log-ratios when listening to a male speaker voice, log-ratios when listening to a female speaker voice ranged closer to zero. This suggests that, independent of adverb gender-stereotypicality and adverb connotation, both robots were looked at more likely when the adverbs were read by a female speaker voice than when being read by male speaker voice.

Figure B13

Log-Transformed Fixation Proportions in the Female Speaker Voice Condition as a Function of Adverb Connotation and Adverb Gender-Stereotypicality by Participants (F_1)



B Experiment 3 – The Effects of Adverb Connotation, Adverb Gender-Stereotypicality, and Speaker Voice on Log-Ratios by Items (F_2)

To test the effect of adverb connotation by items (F_2), a repeated measures MANOVA was performed with log-ratios calculated by items as a function of adverb connotation and adverb gender-stereotypicality as between-items factors and with speaker voice as a within-items factor.

The main effect of adverb gender-stereotypicality was statistically significant, $F_2(1,60) = 4.19$, $p = .045$, $\eta_p^2 = .065$. The main effects of adverb connotation, $F_2(1,60) = 0.96$, $p = .331$, $\eta_p^2 = .016$., and speaker voice, sphericity assumed: $F_2(1,60) = 0.09$, $p = .762$, $\eta_p^2 = .002$, were however not statistically significant. That is, in analyses conducted by items, the adverbs' gender-stereotypicality, but not adverb connotation and speaker voice determined whether the male or the female robot was looked at.

Regarding the interaction effects, the interaction between adverb connotation and speaker voice was statistically significant, sphericity assumed: $F_2(1,60) = 4.85$, $p = .031$, $\eta_p^2 = .075$. This indicates that the adverb connotation dependent of the speaker voice determined whether the male or the female robot was looked at.

The interaction between adverb connotation and adverb gender-stereotypicality, $F_2(1,60) = 0.64$, $p = .428$, $\eta_p^2 = .010$, and between adverb gender-stereotypicality and speaker voice, sphericity assumed: $F_2(1,60) = 0.06$, $p = .806$, $\eta_p^2 = .001$, were not statistically significant. The same was true for the triple interaction between adverb connotation, adverb gender-stereotypicality, and speaker voice, sphericity assumed: $F_2(1,60) < 0.01$, $p = .977$, $\eta_p^2 < .001$. Figure B14 and Figure B15 visualize log-ratios by items as a function of adverb connotation and adverb gender-stereotypically per male and female speaker voice.

Figure B14 illustrates log-ratios by items for the male speaker voice: Reflecting the statistically significant main effect of adverb gender-stereotypicality, the robot whose gender matched (vs. mismatched) the adverbs' gender-stereotypicality was more frequently looked at. In line with the fixation pattern for the male speaker voice by participants, log-ratios were positive, except for stereotypically female negatively connoted adverbs. That is, when the adverbs were uttered by a male speaker, the male robot was more likely fixated than the female one, except when stereotypically female negatively connoted adverbs were uttered. The preference for the male robot was most apparent for stereotypically male positively connoted adverbs.

Figure B14

Log-Transformed Fixation Proportions in the Male Speaker Voice Condition as a Function of Adverb Connotation and Adverb Gender-Stereotypicality by Items (F_2)

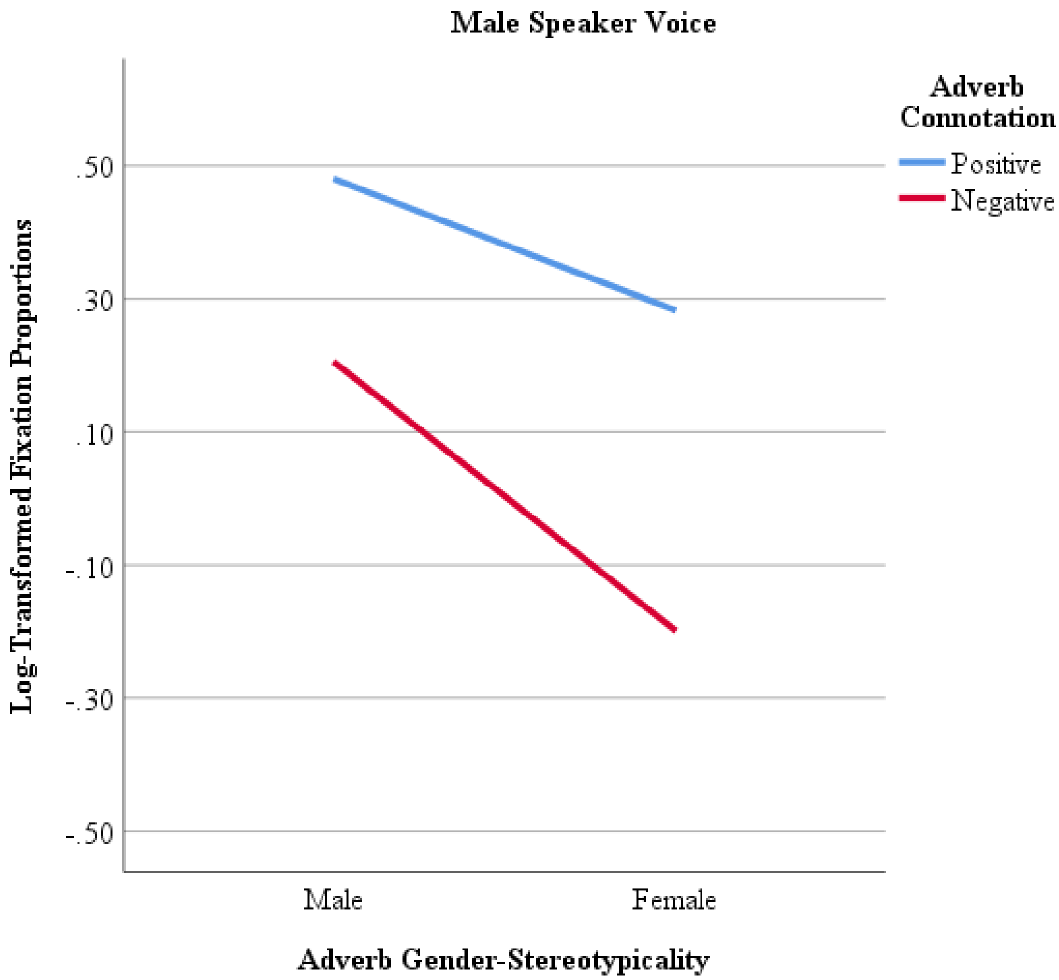
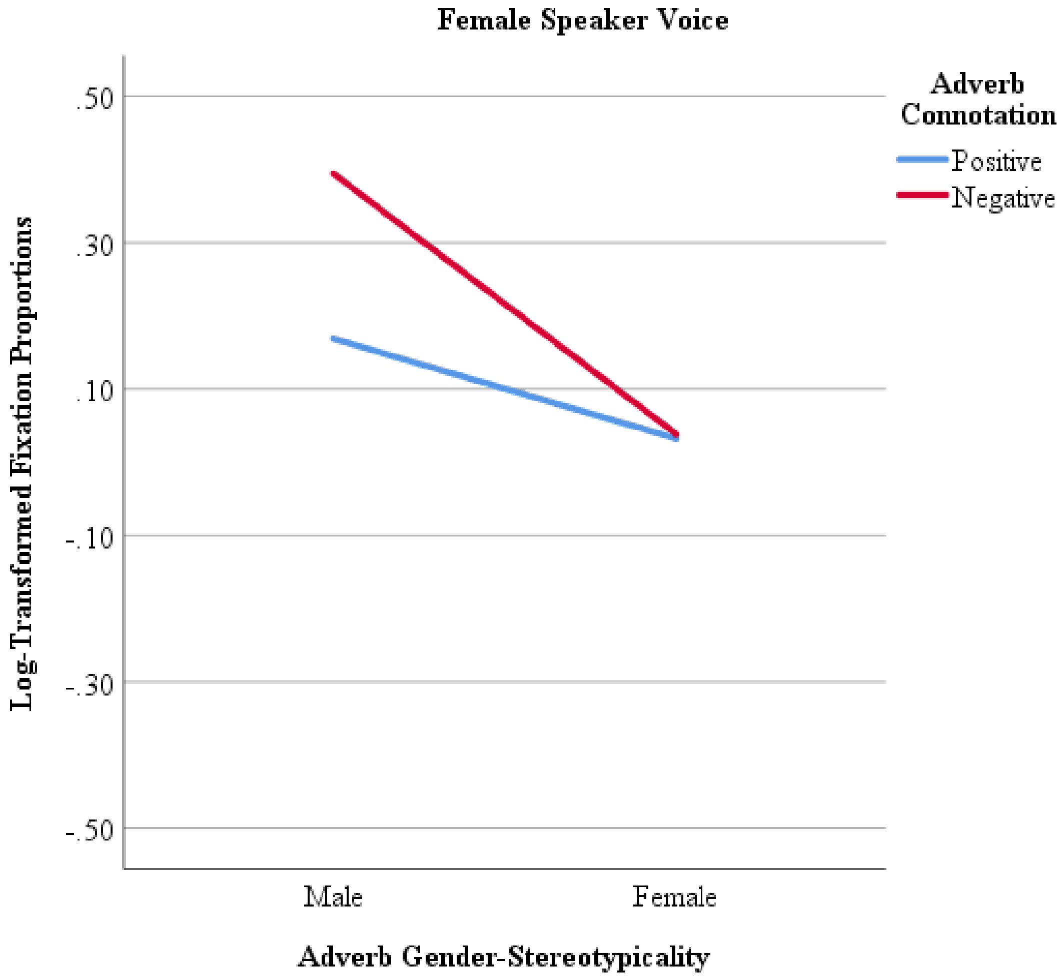


Figure B15 visualizes log-ratios by items for the female voice condition as a function of adverb gender-stereotypicality and adverb connotation: The fixation pattern for the female speaker voice found by items was similar to that found by participants. The robot whose gender matched (vs. mismatched) the adverbs' gender-stereotypicality was more frequently looked at. Positive log-ratios however indicate that the male robot was preferentially fixated over the female one, independent of adverb connotation and adverb gender-stereotypicality. When the adverbs were read by a female speaker, the preference for the male robot was most apparent for stereotypically male negatively connoted adverbs. It was less apparent for stereotypically female adverbs, independent of the adverbs' connotation. Note that, reversely, for the male speaker voice, the preference for the male robot was most apparent for stereotypically male positively connoted adverbs and weakest for stereotypically female negatively connoted adverbs. This might reflect the interaction between adverb connotation and speaker voice that turned out statistically significant by items, but not by-participants.

The fact that the main effect of adverb gender-stereotypicality and the interaction effect between adverb connotation and speaker voice were statistically significant by items, but not by participants suggests that variance across participants, such as due to participants' attitudes or gender, might have played a role. This however was not confirmed[43].

Figure B15

Log-Transformed Fixation Proportions in the Female Speaker Voice Condition as a Function of Adverb Connotation and Adverb gender-Stereotypicality by items (F_2)



[43] In an exploratory analysis, participants' log-ratios were analyzed as a function of adverb connotation, adverb-gender-stereotypicality, and speaker voice, while participants' perceptions of the robot stimuli, their endorsement of the self-report measures, and participant gender were considered as covariates. No considerable effects of the covariates were found on participants' log-ratios for positively and negatively connoted adverbs. Because these analyses would exceed the scope of the present research, they were not reported in further detail.

B Experiment 3 – The Effects of Participants' Perceptions of the Presented Stimuli and Participant Gender on Log-Ratios Within the Adverb Region

Participants' perception of robots as humanlike vs. machinelike entities, as communal and agentic, participants' tendency to anthropomorphize non-human entities, and their ability to recognize the robots' gender and profession, and to imagine robots to perform the presented tasks and professions were deemed important to link the sentences to the robot targets. Participant gender might have affected participants' visual attention when listening to gender-stereotypical adverbs that refer to gendered robots. Therefore, I explored whether participants' perceptions of the robot stimuli and participant gender had guided their visual attention. Particularly the interaction between adverb gender-stereotypicality and external motivation to control for sexist responses had affected participants' visual attention within the adverb region (see Section 13.3.3). Therefore, the effects of participants' internal and external motivation to control for sexist responses and social desirability were also taken into account.

More precisely, a repeated measures MANCOVA was performed with log-ratios calculated by participants (F_1) as a function of adverb gender-stereotypicality as a within-participants factor and speaker voice and participant gender as between-participants factors. The fact that the sample was balanced in terms of participant gender (male voice: n = 18 male, n = 20 female; female voice: n = 17 male, n = 17 female, $\chi^2(1)$= .050, p = .824), allowed to consider participant gender as a between-participants factor (male gender = 0, female gender = 1). This way, the interaction effects between participant gender each with adverb gender-stereotypicality, speaker voice, and the covariates could be investigated.

Participants' levels of robot human- and machinelikeness, agency, communion, tendency to anthropomorphize non-human agents, imaginability of robots to perform the presented *tasks* and *professions*, and the robot stimuli's recognizability and typicality, male and female robot gender, and participants' endorsement of internal and external motivation to control for sexist responses, and social desirability were considered as covariates.

The main effect of adverb gender-stereotypicality was statistically significant. However, the main effects of speaker voice, participant gender, and the covariates were not statistically significant. For clarity of presentation, the results are reported in more detail in Table B11.

Table B11

Main Effects of the Adverb Gender-Stereotypicality, Speaker Voice, Participant Gender, and the Covariates on Participants' Log-Ratios Within the Adverb Region

Covariate	*F*	*p*	η_p^2
Constant	0.08	.777	.002
Adverb Gender-Stereotypicality	8.31	.006	.138
Speaker Voice	2.66	.109	.049
Communion	2.46	.123	.045
External MCSR	2.32	.134	.043
IDAQ	1.70	.198	.032
Imaginability of Tasks	0.93	.339	.018
Female Robot Gender	0.79	.378	.015
Robot Professions' Typicality	0.62	.437	.012
Robot Professions' Recognizability	0.40	.531	.008
Internal MCSR	0.39	.536	.007
Agency	0.13	.718	.003
Social Desirability	0.10	.752	.002
Participant Gender	0.05	.830	.001
Imaginability of Professions	0.05	.833	.001
Robot Humanlikeness	0.02	.904	< .001
Robot Machinelikeness	0.01	.917	< .001
Male Robot Gender	< 0.01	.980	< .001

Note. df(1,52) for all main effects. MCSR = Motivation to Control for Sexist Responses, IDAQ = Individual Differences in Anthropomorphism Questionnaire.

Regarding the interaction effects, the interactions between adverb gender-stereotypicality each with participants' ability to imagine robots to perform the presented *tasks*, external motivation to control for sexist responses, ability to imagine robots to perform the presented *professions*, and social desirability were statistically significant (see Table B12).

Pearson correlations were calculated to explore the direction of the respective interaction effects. They uncovered, the higher participants' endorsement of external motivation to control for sexist responses, $r(70) = -.19$, $p = .114$, social desirability, $r(70) = -.13$, $p = .263$, and the more they could imagine robots to perform the presented *tasks*, $r(69) = -.25$, $p = .039$, and *professions,* $r(69) = -.03$, $p = .825$, the more they looked at the female robot when listening to stereotypically male adverbs. When listening to stereotypically female adverbs, participants looked more at the male robot, the higher their levels of external motivation to control for sexist responses, $r(70) = .28$, $p = .018$, social desirability, $r(70) = .17$, $p = .158$, and the more they could imagine robots to perform the presented *tasks*, $r(69) = .07$, $p = .540$, and *professions*, $r(69) < .01$, $p = .994$.

Table B12

Interaction Effects Between the Experimental Factors and Between Adverb Gender-Stereotypicality and Each of the Covariates on Log-Ratios Within the Adverb Region

Adverb Gender-Stereotypicality x Covariate	*F*	*p*	η_p^2
… Imaginability of Tasks	10.46	.002	.168
… External MCSR	8.85	.004	.145
… Imaginability of Professions	4.26	.044	.076
… Social Desirability	4.19	.046	.075
… Robot Humanlikeness	3.32	.074	.060
… Robot Machinelikeness	2.32	.134	.043
… Robot Professions’ Recognizability	2.27	.138	.042
… Female Robot Gender	1.85	.180	.034
… Participant Gender	1.61	.210	.030
… Communion	1.40	.242	.026
… IDAQ	1.34	.253	.025
… Agency	0.23	.633	.004
… Male Robot Gender	0.20	.655	.004
… Speaker Voice	0.15	.703	.003
… Robot Professions’ Typicality	0.10	.759	.002
… Internal MCSR	0.08	.783	.001
Participant Gender x Speaker Voice	0.44	.511	.008
Adverb Gender-Stereotypicality x Speaker Voice x Participant Gender	0.38	.539	.007

Note. df(1,52) for all interaction effects. MCSR = Motivation to Control for Sexist Responses, IDAQ = Individual Differences in Anthropomorphism Questionnaire.

Following the principle of parsimony (see Tabachick & Fidell, 2007), a similar MANCOVA was performed again, but with participant gender not included and only with participants’ external motivation to control for sexist responses, social desirability, and participants’ ability to imagine robots to perform the presented *tasks* and *professions* as covariates.

The statistically significant main effect of adverb gender-stereotypicality, sphericity assumed: $F_1(1,65) = 11.43$, $p = .001$, $\eta_p^2 = .150$, was confirmed. Moreover, the interaction effects between adverb gender-stereotypicality and external motivation to control for sexist responses, sphericity assumed: $F_1(1,65) = 9.91$, $p = .002$, $\eta_p^2 = .132$, and between adverb gender-stereotypicality and the imaginability of *tasks*, sphericity assumed: $F_1(1,65) = 5.72$, $p = .020$, $\eta_p^2 = .081$, were confirmed. The interaction effects between adverb gender-stereotypicality and imaginability of *professions*, sphericity assumed: $F_1(1,65) = 3.96$, $p = .051$, $\eta_p^2 = .057$, and between adverb gender-stereotypicality and social desirability, sphericity assumed: $F_1(1,65) = 3.16$, $p = .080$, $\eta_p^2 = .046$, turned out not statistically significant.

That is, the adverbs' gender-stereotypicality and the interactions between adverb gender-stereotypicality and each participants' external motivation to control for sexist responses and their ability to imagine robots to perform the presented *tasks* determined whether the male or the female robot was looked at when listening to the adverbs.

B Experiment 3 – The Effects of Participants' Perceptions of the Presented Stimuli and Participant Gender on Log-Ratios Within the NP2 Region

It might have been unexpected for participants to encounter male and female robot targets in Experiment 3. Therefore, it was also explored what additional factors amongst NP2 gender and speaker voice might have affected participants' visual attention when the sentence's target was specified at the end of the sentence.

To do so, analogous to analyses of the adverb region, a MANCOVA was performed with log-ratios calculated by participants as a function of NP2 gender as a within-participants factor and with speaker voice and participant gender as between-participants factors.

Participants' levels of robot human- and machinelikeness, agency, communion, tendency to anthropomorphize non-human agents, imaginability of robots to perform the presented *tasks* and *professions*, and the robot stimuli's recognizability and typicality, male and female robot gender, and participants' endorsement of internal and external motivation to control for sexist responses, and social desirability were considered as covariates.

The main effect of participant gender was statistically significant, while the main effects of NP2 gender, speaker voice, and any of the covariates were not statistically significant. For clarity of presentation, the results are reported in more detail in Table B13.

Table B13

Main Effects of NP2 Gender, Speaker Voice, Participant Gender, and the Covariates on Participants' Log-Ratios Within the NP2 Region

Covariate	*F*	*p*	η_p^2
Constant	1.57	.215	.029
Participant Gender	4.25	.044	.075
Internal MCSR	3.44	.069	.062
Speaker Voice	2.73	.105	.05
Robot Professions' Recognizability	2.23	.142	.041
Robot Professions' Typicality	1.60	.212	.03
Imaginability of Tasks	0.98	.328	.018
Imaginability of Professions	0.66	.422	.012
Male Robot Gender	0.65	.424	.012
IDAQ	0.48	.493	.009
Agency	0.39	.533	.008
Social Desirability	0.37	.545	.007
NP2 Gender	0.33	.569	.006
Female Robot Gender	0.23	.631	.004
Communion	0.09	.766	.002
Robot Humanlikeness	0.06	.812	.001
Robot Machinelikeness	0.05	.825	.001
External MCSR	0.01	.935	< .001

Note. df(1,52) for all main effects. MCSR = Motivation to Control for Sexist Responses, IDAQ = Individual Differences in Anthropomorphism Questionnaire.

Likewise, none of the interaction effects between NP2 gender and speaker voice and NP2 gender and the covariates was statistically significant, while the interaction effect between NP2 gender and participant gender was just not statistically significant (see Table B14).

Table B14

Interaction Effects Between the Experimental Factors and Between Adverb Gender-Stereotypicality and Each of the Covariates on Log-Ratios Within the NP2 Region

NP2 Gender x Covariate	*F*	*p*	η_p^2
… Participant Gender	3.98	.051	.071
… Robot Stimuli's Recognizability	3.58	.064	.064
… Social Desirability	2.72	.105	.050
… IDAQ	1.03	.314	.019
… Robot Humanlikeness	0.99	.325	.019
… Speaker Voice	0.99	.325	.019
… Female Robot Gender	0.41	.525	.008
… Imaginability of Tasks	0.37	.544	.007
… Internal MCSR	0.25	.622	.005
… Male Robot Gender	0.16	.690	.003
… Communion	0.14	.711	.003
… External MCSR	0.14	.714	.003
… Robot Professions' Typicality	0.13	.720	.002
… Robot Machinelikeness	0.03	.872	.001
… Competence	0.01	.914	< .001
… Imaginability of Professions	0.01	.932	< .001
Speaker Voice x Participant Gender	1.63	.208	.030
NP2 Gender x Speaker Voice x Participant Gender	0.06	.815	.001

Note. df(1,52) for all main effects. MCSR = Motivation to Control for Sexist Responses, IDAQ = Individual Differences in Anthropomorphism Questionnaire.

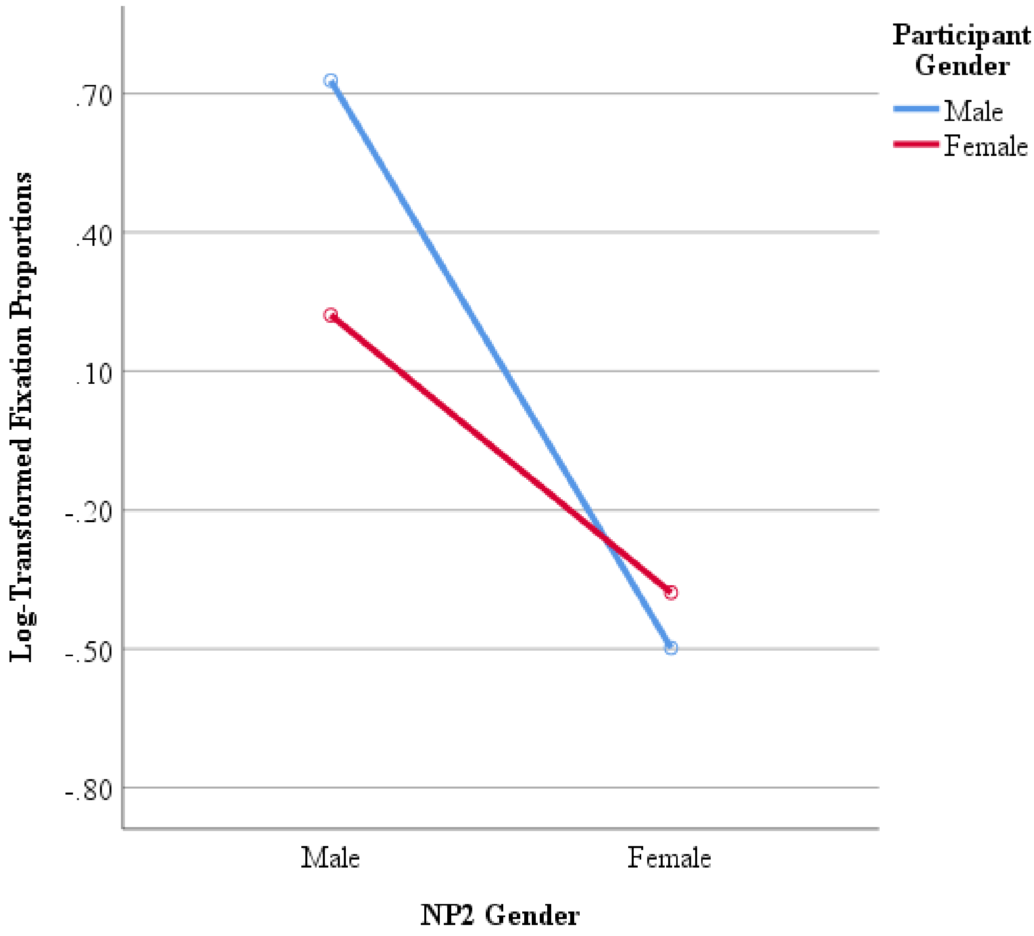
Following the principle of parsimony (Tabachick & Fidell, 2007), a similar analysis was performed with participants' log-ratios as a function of NP2 gender, speaker voice, and participant gender with no covariates included.

The main effect of NP2 gender turned out statistically significant, sphericity assumed: $F_1(1,68) = 45.83$, $p < .001$, $\eta_p^2 = .403$, while the main effect of participant gender did not, sphericity assumed: $F_1(1,68) = 2.97$, $p = .089$, $\eta_p^2 = .042$. However, the interaction between NP2 gender and participant gender, sphericity assumed: $F_1(1,68) = 4.31$, $p = .042$, $\eta_p^2 = .060$, was statistically significant.

To explore the interaction between NP2 gender and participant gender in more detail, participants' log-ratios were displayed as a function of NP2 gender and participant gender in Figure B16: As the statistically significant main effect of NP2 gender suggests, male and female participants both looked at the NP2 target when its gender was specified. Reflecting the statistically significant interaction between NP2 gender and participant gender, male participants looked more frequently at the NP2 target when it was named than female participants.

Figure B16

Participants' Log-Transformed Fixation Proportions as a Function of NP2 Gender and Participant Gender



B Experiment 3 – The Effects of Internal and External Motivation to Control for Sexist Responses, Social Desirability, and Participant Gender on Participants' Responses on the Self-Report Measures

Participants' endorsement of internal and external motivation to control for sexist responses, social desirability, and participant gender might have affected participants' responses on the remaining self-report measures. To test this assumption, a MANCOVA was performed. Participants' mean scores on benevolent and hostile sexism, normative gender role orientation, robot anxiety, robot acceptance, agency, communion, robot humanlikeness, robot machinelikeness, acceptance of technology, competence using technology, control over technology, tendency to anthropomorphize non-human entities, imaginability of *tasks* and *professions*, and participants' evaluations of the robot stimuli in terms of *recognizability*, *typicality*, and *robot gender* were considered as a function of participants' endorsement of external and internal motivation to control for sexist responses, social desirability, and participant gender (male gender = 0, female gender = 1). The results are reported in Table B15. All main effects could not be reported because the table would have been too long and confusing. For a clear presentation of the results, thus only main effects up to $p = .10$ were reported. Pearson correlations were calculated to illustrate the direction of possible effects. Pearson correlations are reported in Table B16.

Table B15

Main Effects of Participants' Endorsement of Internal and External Motivation to Control for Sexist Responses, Social Desirability, and Participant Gender on the Self-Report Measures

Independent Measure	Dependent Measure	F	p	η_p^2
Internal MCSR	Hostile Sexism	33.63	< .001	.341
	NGRO	28.40	< .001	.304
	Benevolent Sexism	20.69	< .001	.241
External MCSR	NGRO	14.14	< .001	.179
	TC Acceptance	6.01	.017	.085
	IDAQ	4.81	.032	.069
	Robot Machinelikeness	4.22	.044	.061
	Robot Humanlikeness	4.14	.046	.060
	Hostile Sexism	4.07	.048	.059
	Imaginability of Professions	3.12	.082	.046
	Communion	2.85	.096	.042
Social Desirability	Benevolent Sexism	9.35	.003	.126
	Robot Acceptance	5.54	.022	.079
	Robot Professions' Recognizability	4.59	.036	.066
Participant Gender	TC Competence	8.06	.006	.110
	TC Acceptance	6.78	.011	.094
	Robot Acceptance	4.57	.036	.066
	TC Control	4.56	.036	.066

Note. $df(1,65)$ for all main effects. NGRO = Normative Gender Role Orientation, MCSR = Motivation to Control for Sexist Responses, TC = Technology Commitment, IDAQ = Individual Differences in Anthropomorphism Questionnaire.

The main effects of participants' internal motivation to control for sexist responses on their responses on hostile sexism, normative gender role orientation, and benevolent sexism were statistically significant. Pearson correlations revealed that the higher participants' internal motivation to control for sexist responses, the lower levels of hostile sexism, $r(70) = -.61$, $p < .001$, normative gender role orientation, $r(70) = -.62$, $p < .001$, and benevolent sexism, $r(70) = -.45$, $p < .001$, they indicated.

The main effects of participants' external motivation to control for sexist responses on their responses on normative gender role orientation, acceptance of technology, tendency to anthropomorphize non-human entities, robot machinelikeness, robot humanlikeness, and hostile sexism were statistically significant. Pearson correlations showed that the higher participants' external motivation to control for sexist responses, the higher their levels of normative gender role orientation, $r(70) = .49$, $p < .001$, acceptance of technology, $r(69) = .36$, $p = .002$, tendency to anthropomorphize non-human entities, $r(69) = .27$, $p = .022$, robot humanlikeness, $r(70) = .26$, $p = .030$, and hostile sexism, $r(70) = .35$, $p = .002$, and the lower their levels of robot machinelikeness, $r(70) = -.21$, $p = .086$.

The main effects of social desirability on participants' responses on benevolent sexism, robot acceptance, and on the robot professions' recognizability were statistically significant. Pearson showed that the higher participants' endorsement of social desirability, the higher levels of benevolent sexism, $r(70) = .18$, $p = .135$, and robot recognizability, $r(70) = .20$, $p = .096$, and the lower levels of robot acceptance, $r(70) = -.30$, $p = .011$, they reported.

The main effects of participant gender on competence in technology use, acceptance of technology, robot acceptance, and perceived control over technology were statistically significant. Female participants indicated lower levels of self-rated competence in technology use, $r(69) = -.31$, $p = .008$, acceptance of technology, $r(69) = -.36$, $p = .002$, robot acceptance, $r(70) = -.28$, $p = .016$, and self-perceived control over technology, $r(69) = -.23$, $p = .053$, than male participants.

Table B16

Bivariate Pearson Correlations Between Participants' Endorsement of Internal and External Motivation to Control for Sexist Responses, Social Desirability, and Participant Gender on the Self-Report Measures

Construct	Internal MCSR	External MCSR	Social Desirability	Participant Gender
Benevolent Sexism	-.45*** [72]	.31** [72]	.18 [72]	-.18 [72]
Communion	.05 [72]	.21† [72]	.02 [72]	-.03 [72]
Hostile Sexism	-.61*** [72]	.35** [72]	-.07 [72]	-.01 [72]
IDAQ	-.06 [71]	.27* [71]	.14 [71]	-.04 [71]
Imaginability of Professions	.01† [71]	.23† [71]	-.11 [71]	-.01 [71]
NGRO	-.62*** [72]	.49*** [72]	-.09 [72]	-.17 [72]
Robot Acceptance	-.20 [72]	.01 [72]	-.30* [72]	-.28* [72]
Robot Humanlikeness	-.01 [72]	.26* [72]	.05 [72]	-.04 [72]
Robot Machinelikeness	.13 [72]	-.20† [72]	-.16 [72]	-.01 [72]
Robot Professions' Recognizability	.14 [72]	-.12 [72]	.20† [72]	0.12 [72]
TC Acceptance	-.23† [71]	.36** [71]	-.16 [71]	-.36** [71]
TC Competence	-.04 [71]	-.11 [71]	.02 [71]	-.31** [71]
TC Control	-.01 [71]	.15 [71]	.04 [71]	-.23† [71]

Note. n indicated in brackets. ***$p < .001$, **$p < .010$, *$p < .05$, †$p < .10$. NGRO = Normative Gender Role Orientation, MCSR = Motivation to Control for Sexist Responses, TC = Technology Commitment, IDAQ = Individual Differences in Anthropomorphism Questionnaire.

B Experiment 4 (Main Verbs - Robots) – Internal Consistencies, Mean Scores, and Pearson Correlations

Table B17

Internal Consistencies (Cronbach's α), Means and Standard Deviations of the Measured Constructs per Speaker Voice

Speaker Voice	Construct	*α*	*M*	*SD*
Male	Benevolent Sexism	.82	2.95	1.04
	Hostile Sexism	.85	2.32	0.83
	NGRO	.80	2.23	0.51
	Social Desirability	.67	4.59	0.65
	Internal MCSR	.82	5.95	0.72
	External MCSR	.77	2.52	0.87
	Robot Anxiety	.87	4.11	1.44
	Robot Acceptance	.88	4.12	1.35
	Communion	.86	2.76	1.20
	Agency	.79	4.58	1.16
	Robot Humanlikeness	.80	2.42	0.88
	Robot Machinelikeness	.79	6.00	0.63
	TC – Acceptance	.83	3.71	1.50
	TC – Competence	.91	5.67	1.36
	TC – Control	.74	4.65	1.05
	IDAQ	.38	1.53	0.51
Female	Benevolent Sexism	.87	3.07	1.27
	Hostile Sexism	.86	2.27	0.93
	NGRO	.88	2.37	0.68
	Social Desirability	.69	4.70	0.74
	Internal MCSR	.80	5.93	0.88
	External MCSR	.82	2.23	0.96
	Robot Anxiety	.81	4.12	1.29
	Robot Acceptance	.88	4.10	1.30
	Communion	.79	3.13	1.24
	Agency	.73	4.70	1.13
	Robot Humanlikeness	.87	2.66	1.03
	Robot Machinelikeness	.81	5.86	0.71
	TC – Acceptance	.86	3.71	1.56
	TC – Competence	.92	5.36	1.56
	TC – Control	.75	5.28	1.01
	IDAQ	.51	1.61	0.60

Note. Male speaker voice: $n = 36$, female speaker voice: $n = 39$. NGRO = Normative Gender Role Orientation, MCSR = Motivation to Control for Sexist Responses, TC = Technology Commitment, IDAQ = Individual Differences in Anthropomorphism Questionnaire.

Table B18

Bivariate Pearson Correlations Between Participants' Log- Ratios for Stereotypically Male vs. Stereotypically Female Main Verbs Referring to Robot Targets and the Covariates

Construct	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1. Log-Ratios for Male Main Verbs																			
2. Log-Ratios for Female Main Verbs	-.18 [71]																		
3. Speaker Voice	-.17 [75]	-.11 [71]																	
4. Participant Gender	-.21† [75]	.04 [71]	-.01 [75]																
5. Benevolent Sexism	.06 [75]	-.11 [71]	.05 [75]	-.30** [75]															
6. Hostile Sexism	.06 [75]	.01 [71]	-.03 [75]	-.03 [75]	.60*** [75]														
7. NGRO	.14 [75]	.03 [71]	.12 [75]	-.16 [75]	.49*** [75]	.69*** [75]													
8. Social Desirability	-.11 [75]	-.11 [71]	.08 [75]	.08 [75]	.17 [75]	.12 [75]	-.06 [75]												
9. Internal MCSR	-.11 [75]	-.08 [71]	-.01 [75]	.23* [75]	-.37** [75]	-.53*** [75]	-.66*** [75]	.10 [75]											
10. External MCSR	<.01 [75]	-.01 [71]	-.16 [75]	-.08 [75]	.21† [75]	.36** [75]	.06 [75]	.02 [75]	-.06 [75]										
11. Robot Anxiety	-.04 [75]	-.05 [71]	<.01 [75]	.02 [75]	.32** [75]	.23* [75]	.24* [75]	.17 [75]	-.13 [75]	.19 [75]									
12. Robot Acceptance	.01 [75]	.07 [71]	-.01 [75]	-.42*** [75]	.05 [75]	.16 [75]	.14 [75]	.01 [75]	-.29* [75]	-.09 [75]	-.25* [75]								

Note. n indicated in brackets. V2s = Main Verbs. NGRO = Normative Gender Role Orientation, MCSR = Motivation to Control for Sexist Responses, TC = Technology Commitment, IDAQ = Individual Differences in Anthropomorphism Questionnaire. Participant Gender/Speaker Voice: 0 = male, 1 = female. Positive log-ratios = male character fixated, negative log-ratios = female character fixated. ****p* < .001, ***p* < .01, **p* < .05, †*p* < .10.

(continued)

Table B18

Bivariate Pearson Correlations Between Participants' Log- Ratios for Stereotypically Male vs. Stereotypically Female Main Verbs Referring to Robot Targets and the Covariates (continued)

Construct	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
13. Communion	-.18	.10	.15	.04	.08	.15	-.02	-.03	-.07	.07	-.19	.26*							
	[75]	[71]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]							
14. Agency	-.10	.07	.05	.24*	.09	.07	-.08	.11	.07	-.12	-.10	.13	.64***						
	[75]	[71]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]						
15. Robot Humanlikeness	-.20†	.22†	.12	.09	.04	.12	-.03	.10	-.06	.11	.09	.10	.72***	.57***					
	[75]	[71]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]					
16. Robot Machinelikeness	-.11	.03	-.10	.25*	-.07	.00	-.33**	.13	.25*	-.16	-.16	.11	.24*	.44***	.22†				
	[75]8	[71]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]				
17. TC – Acceptance	.08	.19	<.01	-.47***	.07	-.03	-.05	.08	-.02	-.06	-.30**	.54***	.25*	.11	.10	.05			
	[75]	[71]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]			
18. TC – Agency	.10	.12	-.11	-.49***	-.03	-.11	-.13	.00	-.05	-.16	-.31**	.39**	.13	-.06	.03	-.06	.51***		
	[75]	[71]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]		
19. TC – Control	-.09	-.14	.30**	-.12	.29*	.03	.16	.26*	-.16	-.13	-.07	.13	.04	.00	-.01	.03	.19	.07	
	[75]	[71]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	
20. IDAQ	-.12	.01	.07	.10	-.04	.04	-.02	.05	.02	.18	.10	.00	.22†	.22†	.33**	-.10	-.01	-.17	-.11
	[75]	[71]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]	[75]

Note. n indicated in brackets. V2s = Main Verbs. NGRO = Normative Gender Role Orientation, MCSR = Motivation to Control for Sexist Responses, TC = Technology Commitment, IDAQ = Individual Differences in Anthropomorphism Questionnaire. Participant Gender/Speaker Voice: 0 = male, 1 = female. Positive log-ratios = male character fixated, negative log-ratios = female character fixated. ****p* < .001, ***p* < .01, **p* < .05, †*p* < .10.

B Experiment 4 (Main Verbs – Robots) – Exploratory Analyses

Exploratory analyses reported in the following were conducted to explain participants' fixation patterns within the main verb and the NP2 region in more detail. To do so, analogous to Experiment 3, participants' ability to recognize the robots' alleged gender and professions and their ability to imagine robots performing the presented tasks and professions were explored. Because participants' perceptions of the robot stimuli might have affected their visual attention, these analyses served to examine participants' fixation patterns within the main verb and the NP2 region in further detail. Moreover, these analyses were done to gain more detailed information about participants' perceptions of robots in general Therefore, analogous to Experiment 3, participants' preferences for robot use, robot acceptance, robot anxiety, and technology commitment were furthermore enquired (see Table B17).

B Experiment 4 – Participants' Ability to Recognize the Robots' Professions and Gender and to Imagine Robots Performing the Presented Tasks and Professions

Similar to Experiment 3, participants' ability to recognize the robots' professions and gender and to imagine robots to perform the presented tasks and professions might have helped them to link the sentences' content to the visual robot stimuli in Experiment 4. Therefore, analogous to Experiment 3, participants' evaluations of the robots' recognizability in terms of profession and male and female robot gender were assessed. Moreover, participants' ability to imagine robots performing the presented tasks and professions were measured using 7-point Likert scales. To quantify participants' evaluations, one-sample *t*-tests were performed against the scale midpoint of 4 (see Table B19)[34]:

In line with in Experiment 3, participants indicated to have recognized the robots' assigned professions which was judged as being portrayed in a typical manner. Within-robot pairs, one robot was clearly perceived as male, while the other was judged as female. Participants' ability to imagine robots to perform the presented tasks and their ability to imagine robots to perform the presented professions were moderate.

Table B19

Results of One-Sample t-Tests Against the Scale Midpoint of 4 on Participants' Evaluations of the Robots' Recognizability in Terms of Profession and Gender, the Professions' Typicality, and Participants' Ability to Imagine Robots to Perform the Presented Tasks and Professions

Item	*M*	*SD*	*t*	*df*	*p*	Cohen's *d*
Recognizability of the Robots' Professions	5.04	1.19	7.57	74	< .001	0.87
Typicality of the Professions' Portraits	4.40	1.54	2.25	74	.028	0.26
Male Robot Gender	5.69	1.33	11.07	74	< .001	1.28
Female Robot Gender	5.65	1.34	10.68	74	< .001	1.23
Imaginability of Tasks	4.11	1.56	0.59	74	.557	0.07
Imaginability of Professions	3.87	1.60	-0.72	74	.472	-0.08

Note. High means indicate high agreement to the respective construct.

B Experiment 4 – Tasks and Professions for Which Robots Would be Used

In addition to participants' ability to imagine robots to perform the presented tasks and professions, I was interested in knowing for what kind of tasks and professions participants would mainly use robots in general. Participants' responses are reported in the following:

***Tasks* Participants Indicated for Which They Would Prefer to Use Robots:** Industrial work, banking, doing hard physical work (e.g., transporting goods), working in dangerous settings (e.g., at height), cleaning, domestic work, repairing things, food delivery, teaching, cooking, sorting things, data management, driving taxi, and handling weapons.

***Professions* Participants Indicated for Which They Would Prefer to Use Robots:** Industrial worker, assistant, secretary, bank clerk, train driver, taxi driver, trucker, craftsman, cleaner, domestic assistant, construction worker, farmer assistant, storekeeper, IT specialist, cashier, assistant in a restaurant (e.g., for taking orders), shop assistant, cook, painter, teacher, mechatronics engineer, technician, and soldier (see also Bernotat & Eyssel, 2018; Bernotat et al.,2021 for similar findings).

To quantify participants' preferences for robot use, participants were asked to rate the task and the profession they had named as *safe* vs. *dangerous*, *interesting* vs. *boring*, *female* vs. *male*, *socially interactive* vs. *socially isolated*, *demanding* vs. *simple*. One-sample *t*-tests against the scale midpoint of 4 were performed (see Table B20)[34]. The results of the one-sample *t*-tests showed that participants would mainly use robots for rather safe[42], boring, and stereotypically male tasks that do not require close human-robot interaction. Regarding participants' evaluations of the professions they had listed, robots were preferred for rather safe and stereotypically male professions.

Table B20

Results of One-Sample t-Tests Against the Scale-Midpoint of 4 on Participants' Evaluations of the Tasks and Professions for Which They Would Use Robots

Category	Item	M	SD	t	df	p	Cohen's d
Tasks	Safe vs. Dangerous	2.80	2.77	-3.76	74	< .001	-0.43
	Interesting vs. Boring	4.84	2.16	3.36	74	.001	0.39
	Stereot. Female vs. Male	4.12	0.43	2.40	74	.019	0.28
	Interactive vs. Isolated	5.35	1.94	6.01	74	< .001	0.69
	Demanding vs. Simple	4.45	2.07	1.90	74	.062	0.22
Professions	Safe vs. Dangerous	3.12	1.79	-4.17	72	< .001	-0.49
	Interesting vs. Boring	4.34	2.10	1.40	72	.167	0.16
	Stereot. Female vs. Male	4.19	0.78	2.11	72	.038	0.25
	Interactive vs. Isolated	4.22	2.08	0.90	72	.372	0.11
	Demanding vs. Simple	3.74	1.99	-1.12	72	.268	-0.13

Note. 7-point semantic differentials were used (e.g., 1 = "safe", 7 = "dangerous"). High means thus indicate high agreement to the latter mentioned construct.

B Experiment 4 – The Effects of Participants' Perceptions of the Presented Stimuli and Participant Gender on Log-Ratios Within the Main Verb Region

Analogous to Experiment 3, it was explored whether participants' perceptions of robots as humanlike entities and of the presented verbal and visual stimuli had guided their eye gazes within the main verb region.

Therefore, similar to Experiment 3, a repeated measures MANCOVA was performed with log-ratios processed by participants (F_1) as a function of main verb gender-stereotypicality as a within-participants factor and speaker voice and participant gender as between-participants factors. The fact that the sample was balanced in terms of participant gender (male voice: n = 18 male, n = 18 female; female voice: n = 20 male, n = 19 female, $\chi^2(1)= .012$, p = .912), allowed to consider participant gender as a between-participants factor (male gender = 0, female gender = 1). This allowed to investigate interaction effects between participant gender each with main verb gender-stereotypicality, speaker voice, and the covariates.

Participants' levels of robot human- and machinelikeness, agency, communion, tendency to anthropomorphize non-human agents, imaginability of robots to perform the presented tasks and professions, and the robot professions' recognizability and typicality, and male and female robot gender were considered as covariates[44].

Only the main effect of speaker voice turned out statistically significant, while the main effects of main verb gender-stereotypicality, participant gender, and any of the covariates were not statistically significant (see Table B21). Similarly, none of the interaction effects was statistically significant (see Table B22). That is, only the speaker voice determined whether participants looked at the male or the female robot.

[44] A similar MANCOVA was performed in which furthermore the effects of participants' endorsement of external and internal motivation to control for sexist responses and social desirability were considered. Because these variables had no statistically significant effects on participants' visual attention within the main verb and the NP2 region, only the more parsimonious analysis was reported without these constructs.

Table B21

Main Effects of Main Verb Gender-Stereotypicality, Speaker Voice, Participant Gender, and the Covariates on Participants' Log-Ratios Within the Main Verb Region

Covariate	*F*	*p*	η_p^2
Constant	0.71	.402	.013
Speaker Voice	5.81	.019	.094
Participant Gender	2.62	.111	.045
Female Robot Gender	2.23	.141	.038
Imaginability of Professions	1.14	.291	.020
Robot Machinelikeness	1.00	.321	.018
Communion	0.96	.331	.017
Male Robot Gender	0.94	.337	.016
IDAQ	0.81	.373	.014
Robot Professions' Recognizability	0.53	.469	.009
Robot Humanlikeness	0.52	.475	.009
Robot Professions' Typicality	0.21	.646	.004
Imaginability of Tasks	0.20	.655	.004
Main Verb Gender-Stereotypicality	0.13	.719	.002
Agency	0.05	.823	.001

Note. *df*(1,56) for all main effects. IDAQ = Individual Differences in Anthropomorphism Questionnaire.

Table B22

Interaction Effects Between the Experimental Factors and Between Main Verb Gender-Stereotypicality and Each of the Covariates on Log-Ratios Within the Main Verb Region

Main Verb Gender-Stereotypicality x Covariate	*F*	*p*	η_p^2
… Participant Gender	2.06	.157	.036
… Robot Humanlikeness	1.92	.172	.033
… Agency	1.11	.296	.019
… Robot Machinelikeness	0.24	.627	.004
... Female Robot Gender	0.23	.632	.004
… Male Robot Gender	0.23	.633	.004
… Imaginability of Professions	0.21	.648	.004
… Robot Professions' Typicality	0.20	.653	.004
… Communion	0.07	.792	.001
… Speaker Voice	0.05	.817	.001
… Robot Professions' Recognizability	0.03	.861	.001
… Imaginability of Tasks	0.01	.907	< .001
… IDAQ	< 0.01	.985	< .001
Speaker Voice x Participant Gender	0.01	.912	< .001
Main Verb Gender-St. x Speaker Voice x Participant Gender	0.12	.736	.002

Note. df(1,56) for all interaction effects. IDAQ = Individual Differences in Anthropomorphism Questionnaire.

B Experiment 4 – The Effects of Participants' Perceptions of the Presented Stimuli and Participant Gender on Log-Ratios Within the NP2 Region

Analogous to the main verb region, it was explored whether participants' perceptions of robots as humanlike entities and of the presented verbal and visual stimuli had guided their eye gazes when the robot target was explicitly named at the end of the sentence.

To do so, a repeated measures MANCOVA was performed with log-ratios calculated by participants as a function of NP2 gender and participant gender as between-participants factors and speaker voice as a within-participants factor.

Participants' levels of robot human- and machinelikeness, agency, communion, tendency to anthropomorphize non-human agents, imaginability of robots to perform the presented tasks and professions, and the robot professions' recognizability and typicality, and male and female robot gender were considered as covariates[44].

The main effect of NP2 gender that was statistically significant when participant gender and the covariates were not considered (see Section 17.4), turned out not statistically significant. The same accounted for the main effects of speaker voice, participant gender, and of any of the covariates (see Table B23).

Table B23

Main Effects of NP2 Gender, Speaker Voice, Participant Gender, and the Covariates on Participants' Log-Ratios Within the NP2 Region

Covariate	*F*	*p*	η_p^2
Constant	0.43	.516	.007
Participant Gender	2.13	.149	.035
Robot Professions' Typicality	1.13	.293	.019
Agency	1.05	.309	.017
Imaginability of Tasks	0.94	.336	.016
IDAQ	0.58	.448	.01
Speaker Voice	0.33	.566	.006
Imaginability of Professions	0.28	.601	.005
Male Robot Gender	0.21	.652	.003
Communion	0.05	.819	.001
NP2 Gender	0.03	.869	< .001
Robot Professions' Recognizability	0.02	.876	< .001
Female Robot Gender	< 0.01	.953	< .001
Robot Machinelikeness	< 0.01	.980	< .001
Robot Humanlikeness	< 0.01	.988	< .001

Note. df(1,59) for all main effects. IDAQ = Individual Differences in Anthropomorphism Questionnaire.

Regarding the interaction effects, only the interaction between NP2 gender and perceived male robot gender was statistically significant (see Table B24). Pearson correlations were calculated to investigate the direction of this interaction effect. They showed that the more participants reported to perceive one of the robots as male, the more they looked at the male robot when listening to a male NP2 target, $r(73) = .36$, $p = .002$, and at the female robot when listening to a female NP2 target, $r(72) = -.21$, $p = .080$.

Table B24

Interaction Effects Between the Experimental Factors and Between NP2 Gender and Each of the Covariates on Log-Ratios Within the NP2 Region

NP2 Gender x Covariate	*F*	*p*	η_p^2
… Male Robot Gender	5.19	.026	.081
… Speaker Voice	1.35	.249	.022
… Robot Machinelikeness	0.99	.323	.017
… Imaginability of Professions	0.78	.381	.013
… Imaginability of Tasks	0.27	.603	.005
… Agency	0.20	.658	.003
… Participant Gender	0.15	.704	.002
… Robot Professions' Typicality	0.13	.718	.002
… Female Robot Gender	0.03	.856	.001
… Robot Professions' Recognizability	0.03	.859	.001
… Robot Humanlikeness	0.03	.876	< .001
… Communion	0.02	.878	< .001
… IDAQ	0.02	.898	< .001
Speaker Voice x Participant Gender	0.51	.477	.009
NP2 Gender x Speaker Voice x Participant Gender	0.09	.763	.002

Note. df(1,59) for all interaction effects. IDAQ = Individual Differences in Anthropomorphism Questionnaire.

Following the principle of parsimony (see Tabachick & Fidell, 2007), a MANCOVA was performed again, with log-ratios by participants as a function of main verbs' gender-stereotypicality and speaker voice and with only participants' perceptions of male robot gender as a covariate.

The main effect of NP2 gender was not statistically significant, sphericity assumed, $F_1(1,71) = 3.49$, $p = .066$, $\eta_p^2 = .047$. The same accounted for the main effects of speaker voice, $F_1(1,71) = 0.51$, $p = .476$, $\eta_p^2 = .007$, and perceived male robot gender, $F_1(1,71) = 1.12$, $p = .295$, $\eta_p^2 = .015$, and for the interaction between NP2 gender and speaker voice, sphericity assumed, $F_1(1,71) = 3.23$, $p = .076$, $\eta_p^2 = .044$. The statistically significant interaction between NP2 gender and perceived male robot gender was confirmed, sphericity assumed, $F_1(1,71) =$

10.01, $p = .002$, $\eta_p^2 = .124$. That is, not the specification of the NP2 target's gender per se, but participants' perceptions of male robot gender dependent of NP2 gender determined whether the male or the female robot was looked at.

B Experiment 4 – The Effects of Internal and External Motivation to Control for Sexist Responses, Social Desirability, and Participant Gender on Participants' Responses on the Self-Report Measures

Similar to Experiment 3, the effects of participants' endorsement of internal and external motivation to control for sexist responses, social desirability, and participant gender might have affected participants' responses on the remaining self-report measures in Experiment 4.

To test this assumption, a MANCOVA was performed with participants' mean scores on benevolent and hostile sexism, normative gender role orientation, robot anxiety, robot acceptance, agency, communion, robot machinelikeness, robot humanlikeness, acceptance of technology, competence using technology, control over technology, tendency to anthropomorphize non-human entities, imaginability of *tasks* and *professions*, and participants' evaluations of the robot stimuli in terms of recognizability, typicality, and perceived male and female robot gender as a function of external and internal motivation to control for sexist responses, social desirability, and participant gender (male gender = 0, female gender = 1). The results are reported in Table B25. All main effects could not be reported because the table would have been too long and confusing. For a clear presentation of the results, thus only main effects up to p = .10 were reported. Pearson correlations were calculated to illustrate the direction of possible effects. For clarity of presentation, Pearson correlations are reported in Table B26.

Table B25

Main Effects of Participants' Endorsement of Internal and External Motivation to Control for Sexist Responses, Social Desirability, and Gender on the Self-Report Measures

Independent Measure	Dependent Measure	F	p	η_p^2
Internal MCSR	NGRO	51.34	< .001	.423
	Hostile Sexism	35.78	< .001	.338
	Benevolent Sexism	9.43	.003	.119
	Robot Acceptance	3.45	.068	.047
External MCSR	Hostile Sexism	13.30	.001	.160
	TC Competence	3.89	.053	.053
	Female Robot Gender	3.78	.056	.051
	Benevolent Sexism	2.80	.099	.038
Social Desirability	TC Control	6.35	.014	.083
	Benevolent Sexism	4.53	.037	.061
	Hostile Sexism	2.99	.088	.041
Participant Gender	TC Competence	25.12	< .001	.264
	TC Acceptance	23.06	< .001	.248
	Robot Acceptance	12.49	.001	.151
	Benevolent Sexism	4.03	.049	.054
	Agency	3.66	.060	.050

Note. df(1,70) for all main effects. MCSR = Motivation to Control for Sexist Responses, IDAQ = Individual Differences in Anthropomorphism Questionnaire.

The main effects of participants' internal motivation to control for sexist responses on normative gender role orientation, hostile sexism, and benevolent sexism were statistically significant. Pearson correlations demonstrated, the higher participants' internal motivation to control for sexist responses, the lower levels of normative gender role orientation, $r(73) = -.66$, $p < .001$, hostile sexism, $r(73) = -.53$, $p < .001$, and benevolent sexism, $r(73) = -.37$, $p = .001$, they indicated (see also Table B26).

The main effect of participants' external motivation to control for sexist responses on hostile sexism was statistically significant. Pearson correlations revealed, the higher participants' external motivation to control for sexist responses, the higher levels of hostile sexism, $r(73) = .36$, $p = .002$, they indicated.

The main effects of social desirability on self-perceived control over technology and benevolent sexism were statistically significant. Pearson correlations showed that the higher participants' social desirability, the higher levels of self-perceived control over technology, $r(73) = .26$, $p = .027$, and benevolent sexism, $r(73) = .18$, $p = .133$, they reported.

The main effects of participant gender on self-perceived competence in technology use, acceptance of technology, robot acceptance, and benevolent sexism were statistically significant. Female participants indicated lower levels of self-perceived competence in technology use, $r(73) = -.49$, $p < .001$, technology acceptance, $r(73) = -.47$, $p < .001$, robot acceptance, $r(73) = -.42$, $p < .001$, and benevolent sexism, $r(73) = -.30$, $p = .009$, than males.

Table B26

Bivariate Pearson Correlations Between Participants' Endorsement of Internal and External Motivation to Control for Sexist Responses, Social Desirability, and Participant Gender on the Self-Report Measures

Construct	Internal MCSR	External MCSR	Social Desirability	Participant Gender
Agency	.08 [75]	-.12 [75]	.11 [75]	.24* [75]
Benevolent Sexism	-.37** [75]	.21† [75]	.18 [75]	-.30** [75]
Female Robot Gender	.07 [75]	-.24* [75]	-.06 [75]	.16 [75]
Hostile Sexism	-.53*** [75]	.36** [75]	.12 [75]	-.03 [75]
NGRO	-.66*** [75]	.06 [75]	-.06 [75]	-.17 [75]
Robot Acceptance	-.29* [75]	-.10 [75]	.01 [75]	-.42*** [75]
TC Acceptance	-.02 [75]	-.06 [75]	.08 [75]	-.47*** [75]
TC Competence	-.05 [75]	-.16 [75]	< .01 [75]	-.49*** [75]
TC Control	-.16 [75]	-.13 [75]	.26* [75]	-.12 [75]

Note. *n* indicated in brackets. ***$p < .001$, **$p < .010$, *$p < .05$, †$p < .10$. NGRO = Normative Gender Role Orientation, MCSR = Motivation to Control for Sexist Responses, TC = Technology Commitment, IDAQ = Individual Differences in Anthropomorphism Questionnaire.

B Post-Hoc Analyses of Statistical Power Per Experiment

Using *G*Power 3.1* (Faul et al.,2007; 2009), post-hoc analyses were performed to calculate statistical power for medium and large effects (see Cohen, 1988; 1992). This was done to better comprehend and interpret the findings of the present experiments (see Stevens, 2009). Analyses were done by participants and by items per experiment by using sample sizes, α-level of .05, and while considering that repeated measures MANOVAs were performed to test the effects of gender-stereotypicality and speaker voice (see Table B27 for the results).

Table B27

Results of Post-Hoc Statistical Power Analyses for Medium and Large Effects (in Parentheses) According to Cohen (1988; 1992) by Participants and by Items per Experiment

		Experiment			
		1	2	3	4
By Participants	Power (1 – β)	.45 (.87)	.41 (.83)	.38 (.80)	.40 (.82)
	Sample Size	86	78	72	75
By Items	Power (1 – β)	.17 (.40)	.16 (.36)	.17 (.40)	.16 (.36)
	Sample Size	32	29	32	29

Note. α = .05. Medium effect: *f* = .25, large effect: *f* = .40 (see Cohen, 1988; 1992).

B Meta-Analysis of All Four Eye Tracking Experiments

The standardized setup of the presented visual world eye tracking experiments (Experiment 1 to Experiment 4) allowed me to perform a meta-analysis and thus to pool the results from all four eye tracking experiments. This in turn yielded an effect estimate which is likely to be more accurate and to have a smaller uncertainty, and thus more statistical power, than estimates of individual experiments (McShane & Böckenholt, 2017). To do the meta-analysis, an online-tool for single-paper-meta-analyses (SPM, McShane & Böckenholt, 2017) was used. The tool was specifically intended for a set of behavioral experiments that share certain similarities (e.g., experimental factors, materials, and samples). Using basic summary information of the experiments and covariances, the tool provided the exact effect estimates and their 50% and 95% confidence intervals (CIs). Moreover, the exact I^2 was provided as a measure of heterogeneity between-experiments (see Higgins & Thompson, 2002; Higgins et al.,2003) as well as its standard deviation, and CIs across experiments. To facilitate future replications of the presented experiments, the SPM tool estimated optimal sample sizes to obtain a level of statistical power of 80% (given $\alpha = .05$) per main effect and per interaction effect across experiments (see supplemental material provided by McShane and Böckenholt, 2017 for detailed information on how effect sizes, heterogeneity, and sample sizes were estimated).

Analogous to hypothesis testing, the main effects of gender-stereotypicality (i.e., adverb and main verb gender-stereotypicality), speaker voice, and the interaction between gender-stereotypicality and speaker voice were estimated across all four eye tracking experiments. To investigate the main effect of gender-stereotypicality, stereotypically male adverbs/main verbs were contrasted against stereotypically female ones. Likewise, to investigate the main effect of speaker voice across experiments, the male speaker voice was contrasted against the female one. Consistent with previous analyses, two meta-analyses were conducted, one by participants and one by items.

B Meta-Analysis on the Effects of Gender-Stereotypicality and Speaker Voice by Participants

The summary information of Experiment 1 to Experiment 4 with data processed by participants (see Table B28), the covariances between male and female gender-stereotypicality within speaker voice conditions (see Table B29), and the contrasts (male vs. female gender-stereotypicality/speaker voice) served as the basis to conduct the SPM.

Table B28

Summary Information of Experiment 1 to Experiment 4 by Participants

Experiment	Gender-Stereotypicality	Speaker Voice	*M*	*SD*	*N*[45]	wi
1 (Adverbs – Humans)	Male Adverbs	Male Voice	-0.31	0.49	43	1
		Female Voice	0.12	0.70	42	2
	Female Adverbs	Male Voice	-0.06	0.60	43	1
		Female Voice	-0.28	0.61	42	2
2 (Main Verbs – Humans)	Male Main Verbs	Male Voice	-0.06	1.01	37	3
		Female Voice	0.06	0.63	41	4
	Female Main Verbs	Male Voice	0.11	1.10	37	3
		Female Voice	-0.17	0.87	41	4
3 (Adverbs – Robots)	Male Adverbs	Male Voice	0.40	0.70	38	5
		Female Voice	0.17	0.75	34	6
	Female Adverbs	Male Voice	0.15	0.97	38	5
		Female Voice	-0.06	0.64	34	6
4 (Main Verbs – Robots)	Male Main Verbs	Male Voice	0.30	1.23	34	7
		Female Voice	-0.05	0.85	37	8
	Female Main Verbs	Male Voice	0.17	0.60	34	7
		Female Voice	0.01	0.78	37	8

Note. By participants, gender-stereotypicality (1. factor) was a within-participants factor, speaker voice (2. factor) was a between-participants factor, wi = sample ID[46].

Table B29

Covariances Between Male and Female Gender-Stereotypicality per Speaker Voice Condition

Experiment	Gender-Stereotypicality	Speaker Voice	Gender-Stereotypicality	Speaker Voice	Cov.
1 (Adverbs – Humans)	Male Adverbs	Male Voice	Female Adverbs	Male Voice	.05
		Female Voice		Female Voice	.01
2 (Main Verbs – Humans)	Male Main Verbs	Male Voice	Female Main Verbs	Male Voice	-.27
		Female Voice		Female Voice	-.26
3 (Adverbs – Robots)	Male Adverbs	Male Voice	Female Adverbs	Male Voice	-.08
		Female Voice		Female Voice	-.07
4 (Main Verbs – Robots)	Male Main Verbs	Male Voice	Female Main Verbs	Male Voice	-.09
		Female Voice		Female Voice	-.21

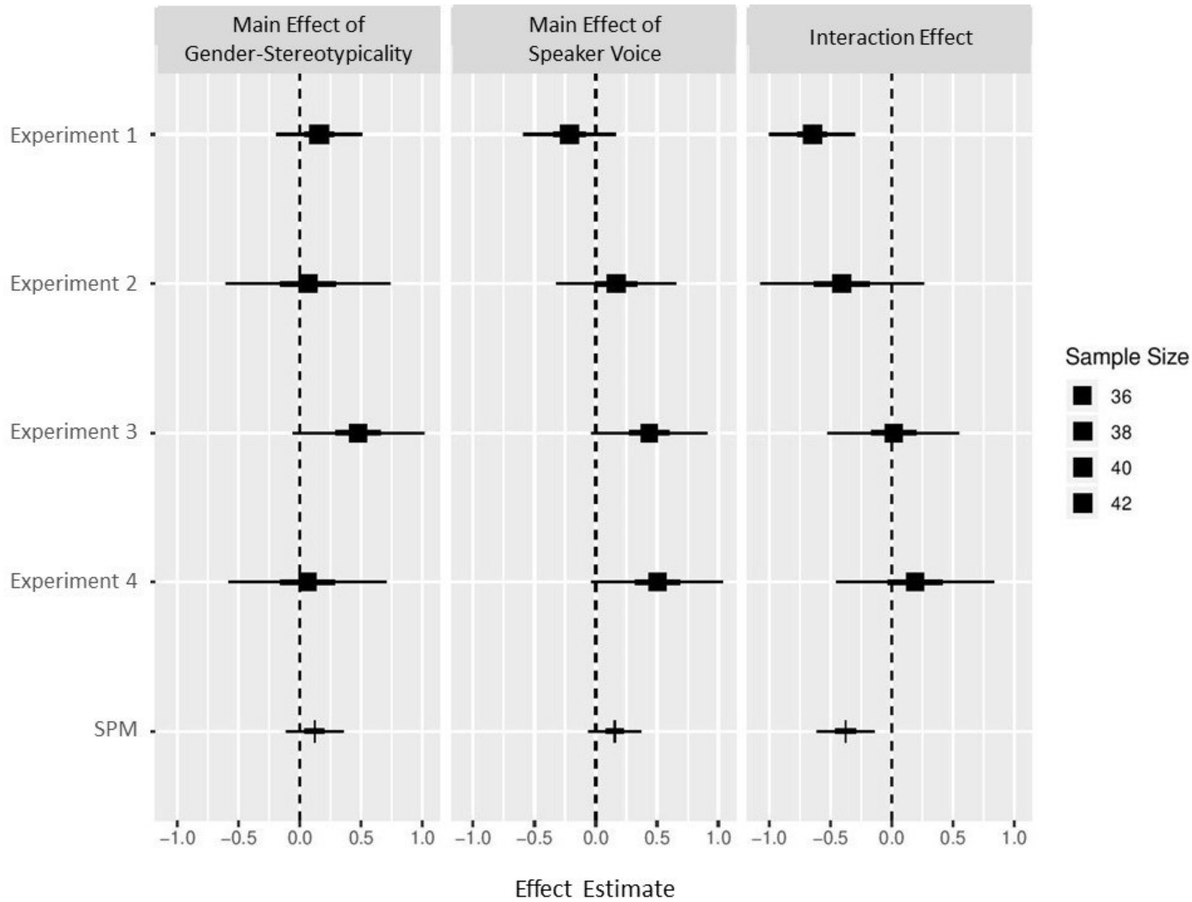
[45] In some cases, *N* had to be lowered to keep sample sizes equal because as a shortcoming of the SPM online-tool, the tool does not accept different sample sizes due to missing data on the within-experimental factor. *N* reported in this section thus corresponds to the data that was inserted into the online-tool.

[46] The sample ID served to enable comparisons between experiments and experimental conditions.

Effect estimates for the main effect of gender-stereotypicality (i.e., adverb and main verb gender-stereotypicality) and speaker voice and for the interaction between gender-stereotypicality and speaker voice per experiment, and overall SPM estimates for each main effect and the interaction between the experimental factors across experiments are displayed in Figure B17. According to McShane and Böckenholt (2017), the figure can be thought of displaying the results of a *t*-test. In case 95% CIs (indicated by the thin lines) overlap the dashed line, an effect can be considered not statistically significant.

Figure B17

Effect Estimates for Each Main and Interaction Effect per Experiment and Overall SPM Estimates for Each Main and Interaction Effect Across Experiments by Participants



Note. Single experiment estimates (indicated by the squares) and the overall effect SPM estimates across experiments (indicated by the vertical bars), and their 50% CIs (indicated by the thick lines) and their 95% CIs (indicated by the thin lines) by participants. The size of the squares indicates the mean sample size per condition in each experiment.

As displayed in Figure B17, the SPM overall estimate for the main effect of gender-stereotypicality was 0.120 (SE = 0.12) which can be considered low. Accordingly, only the 50% CI of the SPM for the main effect of gender-stereotypicality did not include zero, while the 95% CI did. Similarly, the SPM overall effect estimate for the main effect of speaker voice was 0.155 (SE = 0.11). The 50% CI of the SPM for the main effect of speaker voice did not include zero, while the 95% did. Likewise, the interaction between gender-stereotypicality and speaker voice was low with an SPM overall effect estimate of -0.375 (SE = 0.12). Remarkably, however, the 50% and the 95% CIs of the SPM for the interaction effect between gender-stereotypicality and speaker voice did not include zero which implies that the interaction effect can be considered statistically significant (see Table B30 for the effect estimates per experimental condition). I^2 as a measure of heterogeneity between-experiments was approximately 69% (I^2 = 69.24, SD = 0.14, CI_{lower} = 45.48, CI_{upper} = 82.64) which can be considered moderate (see Higgins & Thompson, 2002). Its CIs however range widely. This value of I^2 suggests that 69% of the variation between experiments was caused by differences in the experimental factors or in sample sizes. Interpreting I^2 it needs to be considered that between experiments either gender-stereotypical adverbs or gender-stereotypical main verbs were presented together with either human targets or with robot targets. Furthermore, the adverbs' connotation was balanced, while connotation was not considered for the main verbs. This might explain the moderate level of heterogeneity. Furthermore, to facilitate future replications of the presented experiments, the SPM tool estimated a sample size of n = 19819 for the main effect of gender-stereotypicality to obtain statistical power of 80% (given α = .05). For the main effect of speaker voice, a sample size of n = 2124 was estimated, while for the interaction effect a sample size of n = 141 was estimated to obtain statistical power of 80% (see Table B31). The estimated sample sizes are far larger than sample sizes in Experiment 1 to Experiment 4 which suggests that sample sizes in the present experiments were too small to obtain statistical power of 80%.

Table B30

Effect Estimates and Their Standard Errors for Gender-Stereotypicality and Speaker Voice per Experimental Condition by Participants

Gender-Stereotypicality	Speaker Voice	Estimate	SE
Male Adverbs/Main Verbs	Male Voice	-0.007	0.09
Male Adverbs/Main Verbs	Male Voice	0.103	0.09
Female Adverbs/Main Verbs	Female Voice	0.120	0.09
Female Adverbs/Main Verbs	Female Voice	-0.145	0.09

Table B31

Estimated Sample Sizes (N) Required to Obtain 80% Statistical Power for Each Effect in a Meta-Analysis of 2, 3, 4, and 5 Experiments With By-Participants Data

Effect	Experiment			
	2	3	4	5
Gender-Stereotypicality	9943	6621	4963	3969
Speaker Voice	1063	709	531	425
Gender-Stereotypicality x Speaker Voice	71	47	36	29

B Meta-Analysis on the Effects of Gender-Stereotypicality and Speaker Voice by Items

Using data processed by items, a meta-analysis was conducted analogously to the meta-analysis across data by participants. The summary information of all four experiments with data processed by items (see Table B32), the covariances between male and female speaker voice per gender-stereotypicality (see Table B33), and the contrasts (male vs. female speaker voice/ gender-stereotypicality) served as the basis to conduct the SPM.

Table B32

Summary Information of Experiment 1 to Experiment 4 by Items

Experiment	Speaker Voice	Gender-Stereotypicality	*M*	*SD*	*N*[45]	wi
1 (Adverbs – Humans)	Male Voice	Male Adverbs	-0.22	0.38	32	1
		Female Adverbs	-0.12	0.45	32	2
	Female Voice	Male Adverbs	-0.01	0.55	32	1
		Female Adverbs	-0.27	0.46	32	2
2 (Main Verbs – Humans)	Male Voice	Male Main Verbs	0.11	0.79	29	3
		Female Main Verbs	-0.15	0.73	29	4
	Female Voice	Male Main Verbs	-0.03	0.43	29	3
		Female Main Verbs	-0.03	0.52	29	4
3 (Adverbs – Robots)	Male Voice	Male Adverbs	0.34	0.64	32	5
		Female Adverbs	0.04	0.74	32	6
	Female Voice	Male Adverbs	0.28	0.86	32	5
		Female Adverbs	0.04	0.53	32	6
4 (Main Verbs – Robots)	Male Voice	Male Main Verbs	0.26	0.64	29	7
		Female Main Verbs	0.15	0.45	29	8
	Female Voice	Male Main Verbs	-0.05	0.39	29	7
		Female Main Verbs	-0.03	0.58	29	8

Note. By items, speaker voice (1. factor) was a within-items factor, gender-stereotypicality (2. factor) was a between-items factor, wi = sample ID[46].

Table B33

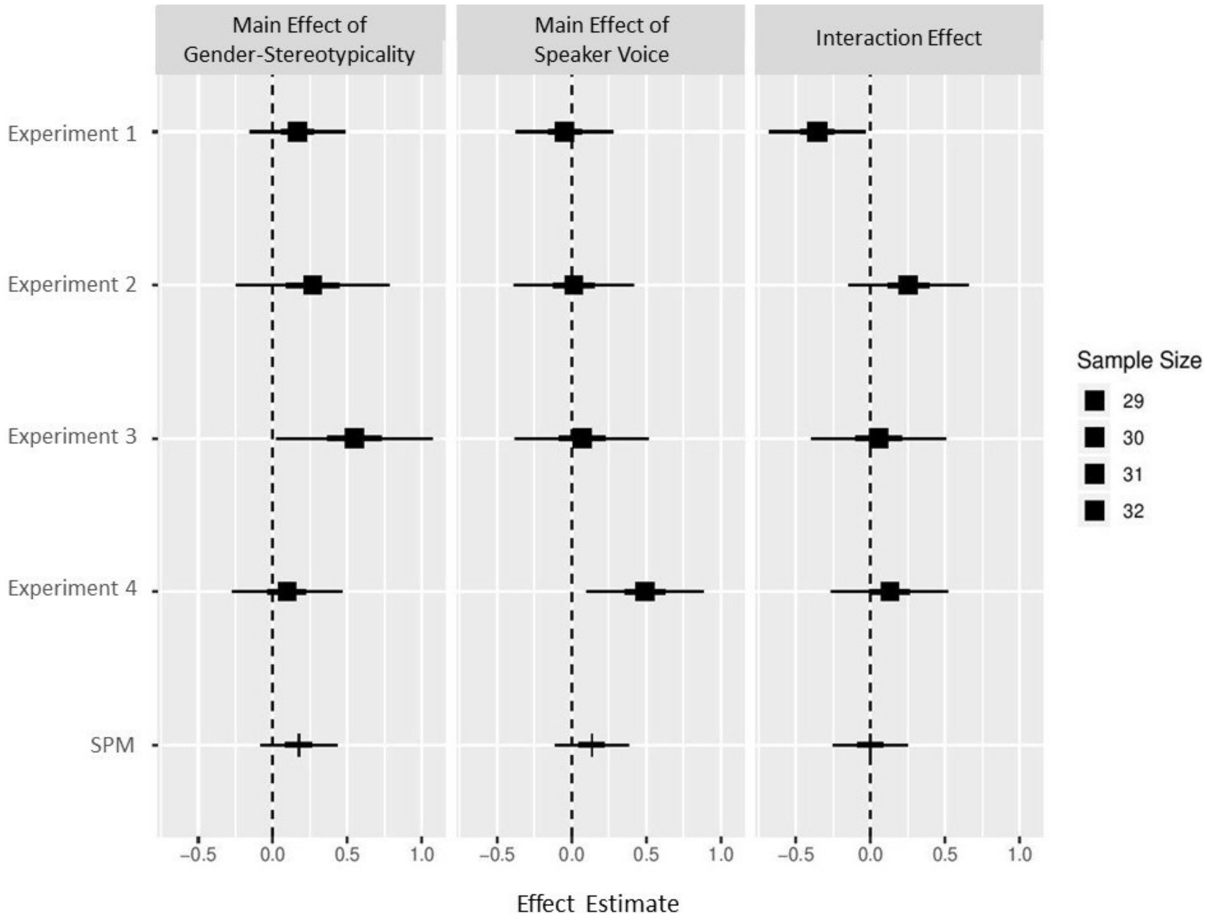
Covariances Between Male and Female Speaker Voice per Gender-Stereotypicality

Experiment	Speaker Voice	Gender-Stereotypicality	Speaker Voice	Gender-Stereotypicality	Cov.
1 (Adverbs – Humans)	Male Voice	Male Adverbs	Female Voice	Male Adverbs	< .01
		Female Adverbs		Female Adverbs	-.01
2 (Main Verbs – Humans)	Male Voice	Male Main Verbs	Female Voice	Male Main Verbs	.04
		Female Main Verbs		Female Main Verbs	.15
3 (Adverbs – Robots)	Male Voice	Male Adverbs	Female Voice	Male Adverbs	.12
		Female Adverbs		Female Adverbs	.03
4 (Main Verbs – Robots)	Male Voice	Male Main Verbs	Female Voice	Male Main Verbs	-.04
		Female Main Verbs		Female Main Verbs	.01

The main effect of gender-stereotypicality (i.e., adverb and main verb gender-stereotypicality) and speaker voice and for the interaction between gender-stereotypicality and speaker voice per experiment, and the overall SPM estimates for each main effect and the interaction between the experimental factors across experiments are illustrated in Figure B18.

Figure B18

Effect Estimates for Each Main and Interaction Effect per Experiment and Overall SPM Estimates for Each Main and Interaction Effect Across Experiments by Items



Note. Single experiment estimates (indicated by the squares) and the overall effect SPM estimates across experiments (indicated by the vertical bars), and their 50% CIs (indicated by the thick lines) and their 95% CIs (indicated by the thin lines) by items. The size of the squares indicates the mean sample size per condition in each experiment.

As illustrated in Figure B18, with an overall SPM effect size estimate of 0.174 (SE = 0.13) the main effect of gender-stereotypicality was relatively small. Accordingly, only the 50% CI of the SPM for the main effect of gender-stereotypicality did not include zero, while its 95% CI did. The same accounted for the main effect of speaker voice with an overall SPM effect estimate of 0.136 (SE = 0.13). Likewise, the interaction between gender-stereotypicality and speaker voice was with an overall SPM effect estimate of -0.001 (SE = 0.13) very small. Accordingly, the 50% and 95% CIs for the SPM of the interaction effect did both include zero which indicates that the interaction effect was not statistically significant across experiments (see Table B34 for effect estimates per experimental condition). With an estimate of I^2 = 71.65 (*SD* = 0.15, CI_{lower} = 50.33, CI_{upper} = 83.28) heterogeneity was still moderate. With regard to future replications of the presented experiments, the SPM tool estimated that 80% statistical power (given α = .05) could not be obtained in a meta-analysis of two to five replications.

To sum up, the overall SPM estimates for the main effects of gender-stereotypicality and speaker voice were small in by-participants and by-items analyses. The same accounted for the interaction effect between gender-stereotypicality and speaker voice. Sample sizes by participants and by items were apparently too small in all experiments to obtain statistical power of 80%. It is thus possible that the effects turned out not statistically significant due to low statistical power. The results of the individual experiments suggest that particularly the adverbs' gender-stereotypicality in interaction with participant gender, participants' sexist attitudes and their intention not to respond gender-stereotypical or sexist may have determined their visual attention during language processing.

Table B34

Effect Estimates and Their Standard Errors for Gender-Stereotypicality and Speaker Voice per Experimental Condition by Items

Speaker Voice	Gender-Stereotypicality	Estimate	SE
Male Voice	Male Adverbs/Main Verbs	0.085	0.09
Male Voice	Male Adverbs/Main Verbs	-0.002	0.09
Female Voice	Female Adverbs/Main Verbs	0.017	0.09
Female Voice	Female Adverbs/Main Verbs	-0.070	0.09

“Man muss Mut haben aufzuhören.“ [You need courage to finish.]

This is what my 8^{th} grade German teacher said to me.

Curriculum Vitae
Jasmin Bernotat
*** 22/02/1985**

Scientific Career & Education

09/2015 – 12/2021	**Doctoral Studies in *Psychology* and *Intelligent Systems* (Dr. rer. nat.)** Center of Excellence Cognitive Interaction Technology (CITEC), Bielefeld University, Germany Thesis: "Keep an eye on stereotypes – The impact of gender-stereotypes (toward robots) on language processing" (eye tracking research)
09/2019 – 02/2020	**Scholar of *Bielefelder Nachwuchsfonds*** [promotion of academic talents] Bielefeld University, Germany
01/2019 – 03/2019	**Research Associate** Social Psychology and Experimental Gender Research, Bielefeld University, Germany
07/2016 – 12/2018	**Research Associate in *Experimentalpsychologische Genderforschung*** [Experimental Psychological Gender Research, third party fund project] Bielefeld University, Germany • **Teaching** and **supervising** B.Sc. and M.Sc. students in Social Psychology, Psycholinguistics, and Gender Research
09/2015 – 12/2018	**Research Associate in *Cognitive Service Robotics Apartment as an Ambient Host (CSRA)*** [third party fund project] Center of Excellence Cognitive Interaction Technology (CITEC), Bielefeld University, Germany • **Coordination** of the Evaluation Group • Research on **Usability** and T**rust in Human-Technology Interaction**
10/2012 – 09/2015	**Master of Science** Psychology, Bielefeld University, Germany
10/2009 – 09/2012	**Bachelor of Science** Psychology, Bielefeld University, Germany
10/2005 – 02/2009	**Nursing Education** Elisabeth Hospital, Damme, Germany
06/2005	**Abitur (German A-Level Equivalent)** Gymnasium Damme, Germany

International Cooperations

07/2017 – Present	Research on the impact of using the “Hugvie” on impression formation and emotions when interacting with an unknown partner, in cooperation with Friederike Eyssel, **Bielefeld University,** Hidenobu Sumioka, and Junya Nakanishi, **ATR, Kyoto & Osaka University, Osaka, Japan**
04/2017 – Present	Research on cognitive and affective trust in HRI, in cooperation with Friederike Eyssel, **Bielefeld University,** Katrin Lohan, **Ostschweizer Fachhochschule, Swiss**, and Muneeb Ahmad, **Swansea University, UK**
01/2020 – 06/2021	Research on the cognitive processes during language comprehension, in cooperation with Gerd Bohner, **Bielefeld University**, Ernesto Guerra, **Universidad de Chile, Santiago**, **Chile,** and Héctor Carvacho, **Pontificia Universidad Católica de Chile, Santiago, Chile**
03/2017 – 03/2020	Research on the impact of robot type on performance and cognition during human-robot interaction, in cooperation with Friederike Eyssel, **Bielefeld University**, Selma Šabanović, Robert Goldstone, **Indiana University, Bloomington, USA,** and Megan Strait, **University of Texas Rio Grande Valley (UTRGV), USA**
03/2017 – 10/2017	Research on the impact of robot body shape on users’ perceptions of robots in Latin America, in cooperation with Friederike Eyssel, **Bielefeld University**, Cesar Lucho, **Pontificia Universidad Católica del Perú, Lima, Peru**, and Gabriele Trovato, **Waseda University, Tokyo, Japan**

Review Activities for International Journals & Conferences

2021	International Journal of Social Robotics
2016 – 2021	International Conference on Human-Robot Interaction (HRI, alt.HRI)
2017 – 2020	International Conference on Robot and Human Interactive Communication (Ro-Man)
2020	International Conference on Human-Agent Interaction (HAI)
2019	Journal of Behavioral Robotics
2019	IEEE Robotics and Automation Magazine
2018	International Conference on Humanoid Robots
2017	Member of the Robot Design Competition Jury (*Softbank Robotics*, *Springer*) at the International Conference of Social Robotics (ICSR)
2016 – 2017	International Conference on Social Robot Interaction (ICSR)

Grants & Awards

03/2020	Best Paper Award of the ACM-IEEE International Conference on Human-Robot Interaction (HRI, Cambridge, UK)
09/2019 – 02/2020	Scholarship of the *Bielefelder Nachwuchsfonds* [promotion of academic talents], Bielefeld University, Germany

02/2019	Best Poster Award for empirical project seminars of the Faculty of Psychology and Sports Sciences, Bielefeld University, Germany
11/2017	Best Paper Finalist Award of the International Conference of Social Robotics (ICSR, Tsukuba, Japan)
06/2017	CITEC-Grant for an *EyeLink 1000 Plus* Eye Tracker Unit (*SR Research*)

Additional Skills

Experimental Software	**Eye-Tracking, Pupillography:** Experiment Builder and Data Viewer (*EyeLink1000, SR Research*), Experiment Center (*SMI*) (very good command) **EMG, EDA, Heart Frequency:** Psychotherapeutic Biofeedback System (*Procom*) (very good command) **Reaction times:** Media Lab (*Empiriesoft*) (basic knowledge) **fMRT:** SPM8 (*Functional Imaging Laboratory* des *Institute for Cognitive Neurology,* London) (fMRT-certificate)
Software for Phonetic Analyses and Annotations	Eudico Linguistic Annotator (*ELAN*) (*Max Planck Institute for Psycholinguistics*), Audacity (*Audacity Team*), Audio Adobe (*Adobe Inc.*) (very good command)
Robot Platforms and Programming	Meka (*Google x*), FloBi, ToBi (*Bielefeld University*), NAO, Pepper, programmed with Choreographe (*Softbank*), (very good command) VDemo, HTML, Python (*Python Software Foundation*) (basic knowledge)
Languages	**German:** native speaker, **English:** fluent (C2-level), **French, Spanish:** good command (B2-level)

Voluntary Work

12/2017	**Co-Founder of the *Coding Community*, Bielefeld University, Germany** Coding group for women
05/2016	**Co-Founder of the *Wissenswerkstatt* at *Impuls Bildungsforum e.V.*,** [private schooling institute], **Bielefeld, Germany** Organization of monthly workshops on health-related topics
10/2014 – 05/2016	**Teacher at the Private Schooling *Impuls Bildungsforum e.V.*,** [private schooling institute], **Bielefeld, Germany** **Courses for students:** German, English, and French (from primary school to A-level), **Courses for refugees:** German, support on every-day issues, and organization of intercultural events

Journal & Proceeding Papers

Ahmad, M. I., **Bernotat, J.**, Lohan, K., & Eyssel, F. (2019). Trust and cognitive load during human-robot interaction. *ArXiv:1909.05160 [Cs]*. http://arxiv.org/abs/1909.05160

Bernotat, J., Eyssel, F., & Sachse, J. (2021). The (fe)male robot: How robot body shape impacts first impressions and trust towards robots. *International Journal of Social Robotics*, *13*, 477–489. https://doi.org/10.1007/s12369-019-00562-7

Bernotat, J., & Eyssel, F. (2018). Can(‘t) wait to have a robot at home? - Japanese and German users’ attitudes toward service robots in smart homes. In J.-J. Cabibihan, F. Mastrogiovanni, A. K. Pandey, S. Rossi, & M. Staffa (Eds.), *Proceedings of the 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (pp. 15–22). IEEE. https://doi.org/10.1109/ROMAN.2018.8525659

Bernotat, J., & Eyssel, F. (2017). An evaluation study of robot designs for smart environments. In B. Mutlu & M. Tscheligi (Eds.), *Proceedings of the 2017 ACM-IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 87-88).

Bernotat, J., Eyssel, F., & Sachse, J. (2017). Shape it – The influence of robot body shape on gender perception in robots. In A. Kheddar, E. Yoshida, S. S. Ge, K. Suzuki, J.-J. Cabibihan, F. Eyssel, & H. He (Eds.), *Social Robotics* (Vol. 10652, pp. 75–84). Springer International Publishing. https://doi.org/10.1007/978-3-319-70022-9_8 **(Best Paper Finalist Award)**

Bernotat, J., & Eyssel, F. (2017). A robot at home – How affect, technology commitment, and personality traits influence user experience in an intelligent robotics apartment. *Proceedings of the 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (pp. 641-646). IEEE.

Bernotat, J., Schiffhauer, B., Eyssel, F., Holthaus, P., Leichsenring, C., Richter, V., Pohling, M., Carlmeyer, B., Köster, N., Meyer zu Borgsen, S., Zorn, R., Engelmann, K. F., Lier, F., Schulz, S., Bröhl, R., Seibel, E., Hellwig, P., Cimiano, P., Kummert, F., … Wrede, S. (2016). Welcome to the future – How naïve users intuitively address an intelligent robotics apartment. In A. Agah, J.-J. Cabibihan, A. M. Howard, M. A. Salichs, & H. He (Eds.), *Social Robotics* (Vol. 9979, pp. 982–992). Springer International Publishing. https://doi.org/10.1007/978-3-319-47437-3_96

Guerra, E., **Bernotat, J.**, Carvacho, H., & Bohner, G. (2021). Ladies first: Gender stereotypes drive anticipatory eye-movements during incremental sentence interpretation. *Frontiers of Psychology,* 12(589429), 1–10. https://doi.org/10.3389/fpsyg.2021.589429

Holthaus, P., Leichsenring, C., **Bernotat, J.**, Richter, V., Pohling, M., Carlmeyer, B., Koster, N., Meyer zu Borgsen, ¨ S., Zorn, R., Schiffhauer, B, Engelmann, K. F., Lier, F., Schulz, S., Cimiano, P., Eyssel, F., Hermann, T., Kummert, F., Schlangen, D., Wachsmuth, S., … Wrede, S. (2016). How to address smart homes with a social robot? A multi-modal corpus of user interactions with an intelligent environment. *Proceedings of the 10th Language Resources and Evaluation Conference of the European Language Resources Association (ELRA)*, 3440-3446.

Lücking P., Lier F., **Bernotat J.**, Wachsmuth S., Šabanović S., & Eyssel F. (2018). Geographically distributed deployment of reproducible HRI experiments in an interdisciplinary research context. *Companion of the 2018 ACM-IEEE International Conference on Human-Robot Interaction Companion (HRI)* (pp. 181-182). Chicago: ACM-IEEE.

Meyer zu Borgsen, S., **Bernotat, J.**, & Wachsmuth, S. (2017). Hand in hand with robots: Differences between experienced and naive users in human-robot handover scenarios. In A. Kheddar, E. Yoshida, S.S. Ge, K. Suzuki, J.J. Cabibihan, F. Eyssel, & H. He (Eds.), *Social Robotics*, *10652*, 587-596. Heidelberg-Berlin: Springer.

Schiffhauer, B., **Bernotat, J.**, Eyssel, F., Bröhl, R., & Adriaans, J. (2016). Let the user decide! User preferences regarding functions, apps, and control modalities of a smart apartment

and a service robot. In A. Agah, J.-J. Cabibihan, A.M. Howard, M.A. Salichs, & H. He (Eds.), *Social Robotics, 9979,* 971-981. Heidelberg- Berlin: Springer.

Strait, M., Lier, F., **Bernotat, J.**, Šabanović, S., Goldstone, R., Eyssel., F., & Wachsmuth, S. (2020). A three-site reproduction of the Joint Simon Effect with the NAO robot. *Proceedings of the 2020 ACM-IEEE International Conference on Human-Robot Interaction (HRI)*, 103-111. **(Best Paper Award)**

Conference Abstracts

Bernotat, J., & Eyssel, F. (2016, September 18–22). *On the tracks of prejudice – Do gender-stereotypical sentences influence gaze directions?* [poster presentation]. 50. Konferenz der Deutschen Gesellschaft für Psychologie (DGPs), Leipzig, Germany.

Bernotat, J., & Eyssel, F. (2017, September 4–6). *What's on a (wo)man's mind? – Der Einfluss des generischen Maskulinums auf mentale Repräsentationen* [What's on a (wo)man's mind? - The influence of the generic masculine on mental representations] [poster presentation]. 16. Tagung der Fachguppe Sozialpsychologie, Ulm, Germany.

Bernotat, J., Eyssel, F., Bohner, G., & Burigo, M. (2019, September 16–18). *Keep an eye on stereotypes – The impact of gender-stereotypical adverbs on language processing* [poster presentation]. 17. Tagung der Fachgruppe Sozialpsychologie, Cologne, Germany.

Bernotat, J., & Sandoval, E.B. (2017, July 14–16). *Let's get ready for the future – A psychological view on the development of an intelligent robotics apartment* [poster presentation]. Junior Expert Workshop of the German-Japanese Exchange Program, Munich, Germany.

Guerra, E., **Bernotat, J.**, & Knoeferle, P. (2015, July 29–30). *Spatial distance between objects affects the interpretation of semantic similarity: evidence from eye movements during spoken sentence processing* [poster presentation]. 8th annual Conference on Embodied and Situated Language Processing (ESLP), Lyon, France.

Schiffhauer, B., Eyssel. F., Kuchenbrandt. D., Bröhl, R., **Bernotat, J.**, & Adriaans, J. (2016). *Erwartungen potentieller Nutzer_innen an ein intelligentes Apartment* [User expectations on an intelligent apartment] [poster presentation]. 50. Konferenz der Deutschen Gesellschaft für Psychologie (DGPs), Leipzig, Germany.

Book Chapters

Lier F., Meyer zu Borgsen S., Wachsmuth S., **Bernotat J.**, Eyssel F., Goldstone R., & Šabanović, S. (2021). Reproducibility in Human-Robot Interaction Research: A Case Study. In P. Cimiano, C. Pietsch, & C. Wiljes (Eds.), *Studies in Analytical Reproducibility: The Conquaire Project* (pp. 129-144). Bielefeld University.

Lectures on Invitation

09/2018 Expert on the topic of "Social Robotics" as part of the "Nerd battle – The science quiz on working fields of the future", "Haus der Wissenschaft", Braunschweig, Germany.

11/2016 Speaker on the topic "New perspectives on the 'we' and 'the others'" as a part of the lecture series "Healthy living", "Wissenswerkstatt", "Impuls-Bildungsforum e.V.", Bielefeld, Germany.

09/2019 **Junior researcher on the topic "**Networks of ideas – Paths and opportunities in Human-Technology Research", **junior researchers and their importance for CITEC research, *CITEC Conference,* CITEC, Bielefeld University, Germany.**