

Pointing and reference reconsidered

Andy Lücking^{a,*}, Thies Pfeiffer^b, Hannes Rieser^c



CrossMark

^a Text Technology Lab, Department of Computer Science and Mathematics, Goethe University, Frankfurt am Main, Germany

^b Excellence Center Cognitive Interaction Technology, Faculty of Technology, Bielefeld University, Germany

^c Collaborative Research Center 673, Faculty of Linguistics and Literary Studies, Bielefeld University, Germany

Received 14 February 2014; received in revised form 20 December 2014; accepted 23 December 2014

Abstract

Current semantic theory on indexical expressions claims that demonstratively used indexicals such as *this* lack a referent-determining meaning but instead rely on an accompanying demonstration act like a pointing gesture. While this view allows to set up a sound logic of demonstratives, the direct-referential role assigned to pointing gestures has never been scrutinized thoroughly in semantics or pragmatics. We investigate the semantics and pragmatics of co-verbal pointing from a foundational perspective combining experiments, statistical investigation, computer simulation and theoretical modeling techniques in a novel manner. We evaluate various referential hypotheses with a corpus of object identification games set up in experiments in which body movement tracking techniques have been extensively used to generate precise pointing measurements. Statistical investigation and computer simulations show that especially distal areas in the pointing domain falsify the semantic direct-referential hypotheses concerning pointing gestures. As an alternative, we propose that reference involving pointing rests on a default inference which we specify using the empirical data. These results raise numerous problems for classical semantics–pragmatics interfaces: we argue for pre-semantic pragmatics in order to account for inferential reference in addition to classical post-semantic Gricean pragmatics.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Keywords: Pointing; Reference; Semantics–pragmatics interface; Computer simulation; Pointing cone; Motion capturing

1. Introduction

The collection of essays on pointing edited by Kita (2003) reveals how complex the interaction of pointing and speech in different contexts of use is. The prototypical pointing gesture, at least in Western cultures, is a bodily behavior which is constituted by an outstretched index finger of one hand. If not stated otherwise, this is also the standard form of pointing we assume throughout this article, although we are aware that there are many ways to perform a pointing act (see for instance Enfield, 2001; Cooperrider and Núñez, 2012). In general, any extended body part or artifact will do the job.

An in-depth reading of the contributions in Kita (2003) shows that little is known about the exact interaction of the pointing gesture with the semantics and pragmatics of the speech it accompanies. It is thus apparent that there is a foundational problem here to be dealt with. Various research traditions (see below) tie pointing closely to reference; hence one might be led to think that unifying pointing and reference research provides a solution to the problem. However, new things are happening in reference research at present, as Herbert Clark and Adrian Bangarter make clear: the assumptions underlying research on reference change, if evidence “from the armchair, laboratory and field” (Clark and Bangarter, 2004:29) is considered. We investigate pointing from the armchair, laboratory and field here. The laboratory and field data we use come

* Corresponding author. Tel.: +49 6979824663.

E-mail addresses: luecking@em.uni-frankfurt.de (A. Lücking), thies.pfeiffer@uni-bielefeld.de (T. Pfeiffer), hannes.rieser@uni-bielefeld.de (H. Rieser).

from experiments and corpus investigations as well as from computer simulation. These will be confronted with the tradition of philosophizing on pointing and reference, and here too, a wealth of new findings is to be considered. This is especially due to the discussion of the semantics–pragmatics interface starting with Relevance Theory, minimalism, and truth-conditional pragmatics (Recanati, 2004a,b). Thus, our investigation is in the spirit of a methodological framework that can be labeled “experimental semantics and pragmatics”. We derive hypotheses from philosophical reasoning about reference and make them an object of empirical falsification. The result of the experimental evaluation is then tied back to theories of grammar. This article, therefore, is in particular of relevance for readers with a background in formal linguistics and analytical philosophy of language as well as for readers with an interest in empirical, methodological approaches to language and its use.

In section 2, we start with a short discussion of pointing from the point of view of various Cognitive Sciences. Based on these results, we then specify semantic and pragmatic hypotheses concerning pointing and reference to be investigated in section 3 in an empirical study which is based on an innovative method relying on motion capturing technology. In section 4, we focus on the empirical study, describing setting, subjects, the annotation techniques and the associated reliability measures. The results of the study are then presented in section 5 where we argue against the semantic hypotheses concerning pointing and reference and opt for pragmatic ones. The discussion is based on various styles of operationalizing the semantic and pragmatic hypotheses involved, above all on the notion of *pointing cone* (cf. Kranstedt et al., 2006a,b,c). In section 6, we discuss the impact of our findings concerning pointing and reference in complex demonstratives for a theory of grammar, especially for its semantics–pragmatics interface. Here, we argue for a revision of classical semantic–pragmatic interfaces extending them with a pre-semantic pragmatics. The next-to-last section (section 7) deals with semantic–pragmatic interfaces that incorporate a non-classic setup from the outset. We pinpoint mere stipulations in Relevance Theory and Situation Semantics and specify how they could be enriched with a mechanism for fixing the situation talked about by plugging in our empirically driven pointing model. Finally, we summarize the argument developed throughout the article in section 8.

2. Cognitive science views on pointing

Linguists claim that pointing is tied up with reference especially with regard to demonstrative pronouns and definite descriptions (Lyons, 1977; Levinson, 1983; Chierchia and McConnell-Ginet, 2000; Haviland, 2000). However, in formal linguistics the relation, if any, between pointings and words used is not investigated in great detail. The relation has been subject to much thought in philosophy, though. We want to emphasize that we cannot attempt to do justice here to the pointing and reference discussion as it has been carried out in philosophy since the Pre-Socratics. Instead, we concentrate on analytical philosophers who relate pointing to semantics, especially to the notion of reference and do so in the most rigid sense possible, e.g. using the tools of modern model theory. As soon will become clear below, this is needed in order to establish the hypotheses for our experimental studies which depend on a strictly defined relation between verbal token and object designated: Peirce (1965) and Wittgenstein (Anscombe and von Wright, 1969) consider pointing as part of the indexical symbol; for a similar point of view see Quine (1960). In the following, we refer to the authors Peirce, Wittgenstein, and Quine as PWQ. Finally, we will rely on the early Wittgenstein, Davidson (1967) and Kaplan (1989a) in order to establish hypotheses which in turn can be tested in our experiments.

At present, three general stances towards (complex) demonstrations show up:

- Farthest away from PWQ is the intentionalism of Kaplan's late work (Kaplan, 1989a). There, demonstration is seen as a mere externalization of intention. While neo-PWQ methodologies exist (see McGinn (1981), an early proponent of the idea that pointing is tied to a vector), there is no neo-PWQian attempt to frame pointings explicitly into some representational format, providing an explicit semantics for them. We observe positions in-between neo-PWQ and pure intentionalism, stressing the contribution of intention and of demonstration for demonstrative reference (Kaplan, 1989b; Braun, 1996, 1994).
- In the Cognitive Sciences, including psychology, research on pointing is frequently related to Kendon's (Kendon, 2000, 2004; Kendon and Versante, 2003) and McNeill's (McNeill, 1992, 2000) work. Extending the pointing notion, McNeill distinguishes pointing into abstract from pointing into concrete domains, defending the thesis that gesture and speech are conceived of as one process, causally linked to one 'growth point'. The 'one process issue' has been investigated in variants of Levelt's model of speech production (Levelt, 1989), for example in work by de Ruiter (2000) or Krauss et al. (2000). Coming from the psychology of social interaction, Clark treats pointing as information on a separate dialog tier (Clark, 1996) and, along with placing, as an attention getter in Clark (2003).
- A research line which has similarities with the one presented below was delineated in Bangerter (2004) and in Bangerter and Oppenheimer (2006), where the detection bias of subjects' pointings was investigated. Clark's attention-centered research on pointing and reference and Bangerter's results on pointing precision and pointing success were brought together in Clark and Bangerter (2004), a paper which unifies attentional and referential research lines concerning pointing.

We draw on all these traditions except on pure intentionalism.

3. Hypotheses concerning pointing

There are two hypotheses about pointing as an ingredient of (borrowing a notion coined in Clark and Bangerter (2004)) *multi-method*¹ communication processes: pointing is either taken to be tied to (direct) reference, especially if it co-occurs with demonstratives, or it is used as an attention getting device, where actual reference has to be brought about by some additional means like verbal description or salience (cf. the distinction made in Kranstedt et al. (2006a)). Accounts taking attention as a general precondition for successful communication like Clark (1996) succeed in harmonizing both these hypotheses. Some researchers treat them as alternatives, though, usually voting for pointing as an attention getter on condition that it does not indicate objects uniquely as for example Butterworth and Itakura (2000). Based on experiments, Bangerter and Oppenheimer (2006) raised serious counter-arguments against the idea that pointing is merely used for attention getting.

Using experiments and corpus investigations, we want to test the claim that pointing is a referring device which means that we first have to look into the context of use of terms like pointing or demonstration in philosophers' theories. Not surprisingly, it turns out that the most interesting contexts to look for are those dealing with names and demonstratives of various kinds. We consult three philosophers concerning pointing or demonstration in this paper, the early Ludwig Wittgenstein, Donald Davidson and David Kaplan, in order to derive hypotheses concerning the function of pointing for reference.

In a manuscript known as *Notes on Logic*, taken down by Bertrand Russell, we find Wittgenstein's dictum "Naming is like pointing" (Anscombe and von Wright, 1969).

Davidson specified the truth conditions for a contextualized sentence like "That book was stolen" in the following way (Davidson, 1967): "That book was stolen" is true as (potentially) spoken by p at t iff² the book *demonstrated* by p at t is stolen prior to t .

Kaplan in turn dealt with the syntax and semantics of demonstratives suggesting that: "[...] each demonstrative, d , will be accompanied by a *demonstration*, δ , thus: $d[\delta]$. The character of a complete demonstrative is given by the semantic rule: In any context c , $d[\delta]$ is a directly referential term that designates *the demonstratum*, if any, of δ in c , and that otherwise designates nothing. Obvious adjustments are to be made to take into account any common noun phrase which accompanies or is built into the demonstrative" (Kaplan, 1989a:572).³ In this paper, we are essentially concerned with providing an explication for Kaplan's δ (the demonstration) and not with referring intentions of agents pointing or with addressees' interpreting pointing acts. It has to be made very clear that the referent of $d[\delta]$ in context c ⁴ is determined by the demonstration δ alone, not by d : What picks out an object is not the (complex) demonstrative, but the demonstration act. Coincidentally, all approaches discussed so far, the empirical as well as the philosophical, deal with *singular reference*, that is, with pointing at single objects instead of plural ones, and so do we. Our methodology introduced below could of course be extended to sets of objects referred to by a pure or a complex demonstrative such as "these" or "these red screws". However, according to our triangular methodology, this would demand an additional empirical study on sets which we leave to future research.

As the references show, some philosophers at least (and not the minor ones), say that pointing is tied up with reference. However, reference can be established at least by two means. On the one hand, an object can be singled out in virtue of being the only thing that fits a certain description. This view has been spelled out by the quantificational account on the level of logical form by Russell (1905). On the other hand, some objects are assumed to be nomologically, most notably causally, linked to some words, as proposed by the causal chain-theory for proper names of Kripke (1980).⁵ Accordingly, semantic accounts of definiteness and reference usually range from neo-Russellian descriptivism to (direct) referentialism.⁶ With that said, we see that all three philosophers tie up pointings with a direct referential mode of operation. In particular:

Wittgenstein indicates that pointing is comparable to naming, names being the prototypical devices to refer, in short: pointing is *the* referring device.

Davidson, providing the truth conditions for a complex demonstrative "that book" uses the notion of demonstration in the meta-language without a corresponding term in the object language. Observe that we have an interaction of definite description and demonstration, but whereas we have a clear understanding of the definite description in the meta-language, presumably due to our object language intuitions concerning definite descriptions, we lack a clear understanding for "demonstrated" (cf. Braun, 1996:146).

¹ Clark and Bangerter (2004) describe how the concept of reference changed due to the investigation of multi-modal data: "When people were videotaped in conversation, they were found to exploit a range of signals that were not linguistic at all. Referring was seen to be a multi-method process, one that normally requires more than one method of signaling."

² "iff" abbreviates "if and only if" in logical parlance.

³ Kaplan's paper is quoted according to the reprinted version of the collection *Themes from Kaplan*.

⁴ Technically, a context c determines a quadrinomial index $\langle c_A, c_T, c_P, c_W \rangle$ consisting, in that order, of the agent of c , the time of c , the position of c , and the world of c (Kaplan (1978a:88); Kaplan (1989a:543)).

⁵ See also many remarks of Peirce on indexical signs, for instance, in his classical paper (Peirce, 1867).

⁶ We are simplifying things a bit, since we remain silent about reference being a property of word or sentence types, or (part of) a specific speech act, acted out in tokens of pragmatic use. But this simplification is harmless to our argument at this point.

The standard theory for demonstrative reference, nevertheless, has been spelled out by Kaplan in his *Demonstratives*. Accordingly, we have a closer look at his analysis subsequently in order to derive hypotheses on pointing gestures carefully. In Kaplan's work, pointing gestures are put into play due to the incompleteness of demonstratives: demonstratives make up the subset of indexical expressions that are incomplete and need to be completed by a *demonstration* δ (see Kaplan, 1989a:490, 524); see also Levinson (2008:101)). The structure ' $d[\delta]$ ' is called a *complete demonstrative* and is a "directly referential term that designates the demonstratum, if any, of δ in [context] c " (Kaplan, 1989a:527). A term is directly referential if, when put in context, it determines a fixed (i.e., stable for all world/time indices) referent as being part of the world and the time of that very context (Kaplan, 1989a:493, 499, 513).

Being directly referential is a shared feature of indexicals (pure and demonstratives alike), and proper names (Kaplan, 1989a:562). However, unlike proper names, the meaning of indexicals is context-dependent, therefore "proper names are not indexicals" (Kaplan, 1989a:562 fn.31). Accordingly, there is no *a priori* reason for treating them on a par (*contra* Wittgenstein) or devising a unified theory comprising both. Rather to the contrary, there seem to be at least three kinds of directly referential expressions acknowledged by Kaplan, namely pure indexicals, demonstratives plus demonstrations, and proper names. Pure indexicals refer to an object x *qua* the conventional meaning rules associated with them (Kaplan, 1989a:505, 532). For instance, the rule "the speaker or writer of the relevant occurrence of the word 'I'" determines the referent of the word "I" in each context of its occurrence.⁷ Proper names are assumed to be bound up with their name-bearer by means of a Kripke-style causal chain (Kaplan, 1989a:563). Demonstrative indexicals are not equipped with such direct referent-providing rules (Kaplan, 1989a:490), although they may provide "sortal" information as is the case, for instance, with pronouns (*he*: male) or complex demonstratives (*that woman*: female) (Kaplan, 1989a:524). Rather, the referent of a proper demonstrative is somehow given by the obligatorily co-occurring demonstration (see above). We say "somehow" since Kaplan, apart from some remarks here and there, remains silent about how demonstrations accomplish reference, as he himself concedes (Kaplan, 1989a fn.9) – we have to come back to this issue shortly.

The theory of demonstration that Kaplan finally subscribes to in *Demonstratives* is the *Corrected Fregean Theory of Demonstratives* (Kaplan, 1989a:528). According to this theory, the referent of a demonstrative is determined by a demonstration relation δ , which is part of the *character* of the demonstrative, that is, a relation from contexts to contents (where contents are functions from circumstances of evaluation, i.e., worlds and times, to extensions) (see also Perry, 2009:188).⁸

Concerning how direct reference actually works in case of demonstrations, Kaplan gives two quite different examples. The first example draws on pointing gestures: being pointed at is taken to be the relevant feature of some object x that allows for identifying x precisely in virtue of that feature (Kaplan, 1989a:513). The second example seems to invoke some kind of salience. Here, an object x is assumed to virtually point at itself *qua* its specific "appearance" seen from a specific perspective (Kaplan, 1989a:525–526). Note that in this case no demonstration act of the speaker is required at all. Since no object used in the demonstration domain of the experiment reported below stands out due to its appearance (say, size, color, . . .) to such an extent that it is self-demonstrating, the speakers in the experiment had to use pointing gestures in order to establish an identifiable relation δ to the respective demonstratum. In case of pointing gestures this relation is fundamentally non-arbitrary in that it is intimately tied to the demonstratum – see the remarks on perspective and demonstration given in footnote 48 of *Demonstratives*. In other words: the referent functions as an object whose features (e.g., its location) partially explain certain features (e.g., orientation) of the demonstration/pointing. That is, Kaplan at this point implicitly draws on a *causal theory of reference* (cf. McGinn, 1981:158) in order to account for the direct referentiality of demonstrations. Understanding this causal relation is in turn part of identifying the demonstratum. What is demonstrated, what can be a *demonstratum*, is not problematic for Kaplan. The notion of concrete object or individual is in itself not called into question by the type of Frege puzzles Kaplan's reasoning deals with. The demonstrative act δ is tied to perception and a perspective. Below it will become clear that we try to explicate these attributes of δ using the pointer's position, the domain or space she is pointing into, the pointing device and variously, a pointing ray or a pointing cone, giving in the end preference to cones.

Turning to Davidson, Davidson's theory is extending Tarski's with parameters that help to reconstruct the meaning of situated utterances leading to a contextualized notion of satisfaction (see Lepore and Ludwig (2007:76–99) for details). However, not being tied to a particular brand of reference theory, it could also be reconciled with a Fregean approach. In contrast, D. Kaplan considers sentences in Context-Character-Content Structures (Perry, 2009:490). Notwithstanding these differences, all theories of demonstration need a working notion of pointing/demonstration in order to cover mundane uses of indexicals. Hence, the question to be asked is "What does it mean to demonstrate by pointing?". This will be one of the issues we will deal with below, using measuring data generated by tokens of pointing acts.

We explicate these views on demonstrations/pointings with the help of psychologically oriented accounts by several operationalizations involving pointing beams and put it to empirical test (see below). As a result, we argue for abandoning

⁷ In order to secure direct referentiality for such meaning rules technically, Kaplan proposes his *dthat* operator as a uniform modeling device (see also Kaplan, 1978b).

⁸ Accounting for the context-dependent meanings of indexicals by means of the notion of characters correctly captures the truth-conditional difference between *I am tall* and *The speaker is tall* (see also Kaplan, 1978a).

the Kaplanian direct referentialist account in favor of a variant of a descriptive theory of demonstrations as argued for by [Geurts \(1997\)](#) by example of proper names (though acknowledging generalizability to other phenomena including demonstrations) and adopted by [Zeevat \(1999\)](#).

So, putting together these philosophers' findings applied to pointing acts, we arrive at the following hypothesis:

Hypothesis 1. (Sem): A demonstration [pointing] going together with a simple or a complex demonstrative in context *c* designates exactly one object, the object referred to in *c*.

Taking into account that there are at least two parties involved in this process, producers and recipients, we will work with two variants of (Sem) of different scope. In an ideal world, one would expect (Sem) to hold for all interlocutors and we formulate this as:

Hypothesis 2. (Inter Sem): (Sem) holds interpersonally for producers and recipients of demonstrations.

However, we acknowledge that timing and differing perspectives of producers and recipients are real-world problems that might interfere with our ideal world. Therefore (Sem), if valid, should at least hold for the producer of the demonstration:

Hypothesis 3. (Intra Sem): (Sem) holds intrapersonally for the producers of demonstrations only.

In the following, we will primarily refer to (Sem) as the base hypothesis and only differentiate between (Inter Sem) and (Intra Sem) when we are interested in the scope of (Sem).

Intuitively, if the philosophers' definitions are not vacuous, there must be some way to operationalize them. Scanning the literature for mechanisms for (Sem), especially the work of Butterworth and peers gets into focus. [Butterworth \(2003:24\)](#), for instance, proposes a near behavioral equivalent to designating as used in (Sem), namely vector extrapolation: we are led to identify demonstration by pointing with a vector anchored in and oriented along the lever used (index finger, pointing stick or whatever). We may conceive of this vector as a model for the pointing ray generated in the pointer's aiming (cf. Bühler's *origo*) at the object he wants to be identified for his audience and likewise being reconstructed in pointing comprehension. In a closely related manner, [McGinn \(1981:163\)](#) suggests that the referent of pointing is the first object that intersects a line projected from the pointing finger. Thus, the vector model provides a reasonable operationalization of the direct referentiality of pointings.

The vector model, however, has been criticized right from the start. Butterworth and Itakura seem to have found experimental evidence against the vector extrapolation hypothesis: for adults, following head and eye cues was at least as accurate as following pointing. And again there was no evidence that pointing comprehension depends on precise linear vector extrapolation ([Butterworth and Itakura, 2000:48](#)). Thus, if there is vector extrapolation it is at best approximate and sufficient only to differentiate between widely spaced, objects. In crowded conditions typical of real life, the differential salience of the target, its motion or its familiarity may help to single it out. On this view, ecological factors necessarily interact with approximate postural cues in establishing joint attention and in making definite reference ([Butterworth and Itakura, 2000:49](#)). These findings come close to [Bangerter \(2004\)](#) and to [Clark and Bangerter \(2004\)](#).

Mainly relying on the non-philosophical literature, we establish two further hypotheses, the first one shifts the emphasis from semantic interpretation to inference to an object and the second one is concerned with the focus of attention, doing away with the notion of an object referred to. Both hypotheses are taken as belonging to the domain of pragmatics. The first one is called "Strong Prag" and the second "Weak Prag".

Hypothesis 4. (Strong Prag): A demonstration is produced to trigger a perceptually based inference with regard to a context *c* from the pointing device to the object referred to in *c*.

Hypothesis 5. (Weak Prag): A demonstration is produced to shift its addressee's attention towards a specific domain in a context *c*.

Our experiments will try to investigate which one of these hypotheses can be substantiated and provide the basis for computer modeling and theory construction. By way of explication, Butterworth and Itakura's findings seem to falsify (Strong Prag) and yield support to (Weak Prag), which in fact seems to be identical with the notion of reference in some psychological theories (e.g. [Pechmann and Deutsch, 1982](#)). They are not related with (Sem), though, which is only concerned with the interpretation of tokens. To substantiate the hypotheses above, we proceed in the following way: if we can show that (Sem) is what characterizes pointing acts, we obviously do not need the pragmatics hypotheses, since rules of symbol use alone can care for reference in co-verbal pointing. So, the first step will be to have a look at that (see sections 5.1 and 5.2). If we find that pointing success is tied to various pragmatic parameters, we have to consider the pragmatics hypotheses (see section 5.4). Obviously, if we can find evidence for (Strong Prag), we will have proved (Weak Prag), granted that we tie (Strong Prag) to intention and attention. Anyway, (Weak Prag) alone would not be of much help, since it is too weak to distinguish pointing from focusing or emphasizing.

The different hypotheses are investigated empirically, using motion capturing ([Pfeiffer et al., 2006](#)) in an experimental setting comprising a Description Giver (producer), an Object Identifier (recipient) and a pointing domain ([Kranstedt et al., 2006b,c](#)); further details are given below. If (Sem) cannot be falsified, we would expect that every one of the Description Giver's multi-modal acts consisting of the description of an object and a co-verbal pointing uniquely designates an object

(Intra Sem) which then could be identified by the Object Identifier (Inter Sem). If neither is the case, we will have to turn to the pragmatics hypotheses and investigate those.

4. Empirical study

In order to test the hypotheses concerning pointing set up above, we conducted an empirical study using an object identification game setting. This setting was developed from an earlier study on the use of pointing gestures in referring (Lücking et al., 2004). We made two changes with respect to the previous experimental scenario. Firstly, the positioning of the objects that make up the pointing domain has been randomized, while the previous study draws on two motivated clusterings, one according to form and one according to color (Kühnlein and Stegmann, 2003). Secondly, the previous study relied exclusively on audio and video data and their annotations. The decisive improvement in the experiment under discussion is the additional use of tracking technology for motion capturing.

4.1. Experimental procedure

The empirical study rests on object identification games, which involve two participants each. Each participant gets a certain role: one is called *Description Giver* (henceforth DG) and the other *Object Identifier* (OI). DG and OI are placed within the operational area of a marker-based optical tracking system with nine cameras (ARTTRACK1 System, ART GmbH). The information delivered by the cameras is integrated via special software and provides positions and orientations of optical markers in an absolute coordinate system. Only the DG is tracked. He sits on a stool and is equipped with carefully positioned markers on his arms, index fingers, hands, and head. The tracking system is clocked by the video frame rate (1/25 s). In addition, the whole scene is recorded from two different perspectives with digital cameras. Speech is captured with the DG's headset. In order to facilitate a synchronized recording of the various data streams, our own software framework, the *Interactive Augmented Data Explorer* (IADE) (Pfeiffer et al., 2006), was used. IADE provides functionality to process as well as to simulate recorded multi-modal data – features we need in order to analyze the tracked pointing behavior of our participants. The whole set-up with the prepared DG can be seen in Fig. 1, a screen-shot from our video recordings. The customized gloves used to track the stretched index finger are displayed in Fig. 2.

Both DG and OI are located around a real table (77.5cm × 155.5 cm) with 32 parts of a Lorentz Baufix toy air-plane, the experimental domain. The outer objects' centers frame an area of 140 cm × 70 cm. The objects' centers were lined up on an underlying grid, see Fig. 3. Note for the sake of upcoming references that the domain is divided up from the perspective of the DG into eight rows along the longer and four columns along the shorter side of the table. The rows are numbered starting from the position of the DG (see Fig. 3).

Each experimental run of an identification game comes in two variations (sub-settings), differing in the communicative channels (speech and gesture) the DG is allowed to use. In the first sub-setting, the DG can use speech as well as gesture (we abbreviate this variation as S+G-Trial). In the second sub-setting, the DG's behavior is restricted to gesture only (G-Trial).

The interaction between the Description Giver, sitting on a stool, and the Object Identifier, placed at the opposite side of the table, is observed by two video cameras and nine cameras of a motion capturing system. The order in which the DG

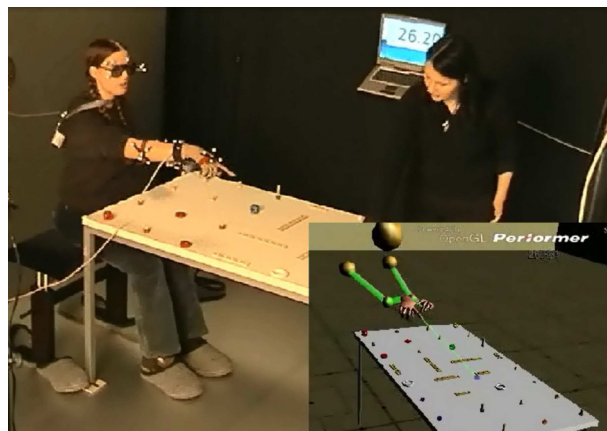


Fig. 1. The experimental setup: the Description Giver sits to the left of the table, the Object Identifier stands to the right (opposite of the DG and not sideways as shown in the picture) and has a pointer (not displayed here). The screen-shot in the lower right corner shows a reconstruction of the posture of the DG and the direction of her pointing gesture from the recorded motion capturing data. The small green ray extending from the visualization of the finger shows the pointing ray extending from the finger tip. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)



Fig. 2. Customized gloves for tracking the index fingers.

has to demonstrate the objects is fixed in a pre-setting map, which is displayed on a monitor in front of him (see M1 in Fig. 4). To abstract from potential sequence effects, six pre-setting maps have been randomly generated.

The interaction between DG and OI is highly restricted to avoid uncontrollable negotiation processes between the interactors. It consists of four formalized steps fixed in the briefing:

1. Demonstration by DG (bimodal or gestural, according to respective sub-setting);
2. Interpretation by the OI;
3. Identification act by OI using only a pointer (the referent remains in its place);
4. Verbal feedback by DG.

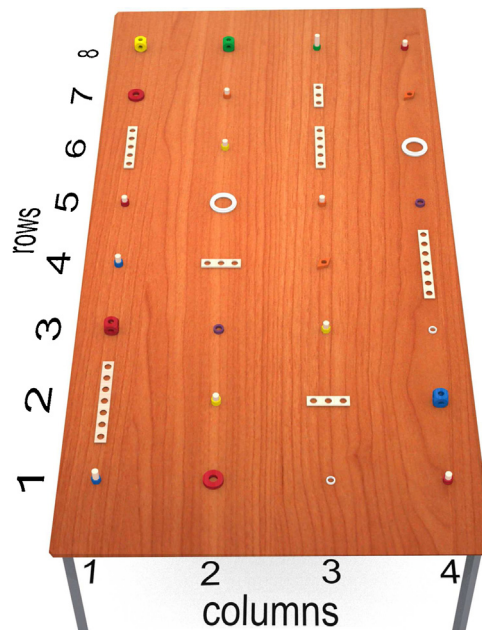


Fig. 3. The experimental domain is divided up from the DG's perspective into eight rows and four columns. It covers an area of 70 cm × 140 cm. The image shows one out of six templates of positioning objects in the domain (see main text for details).

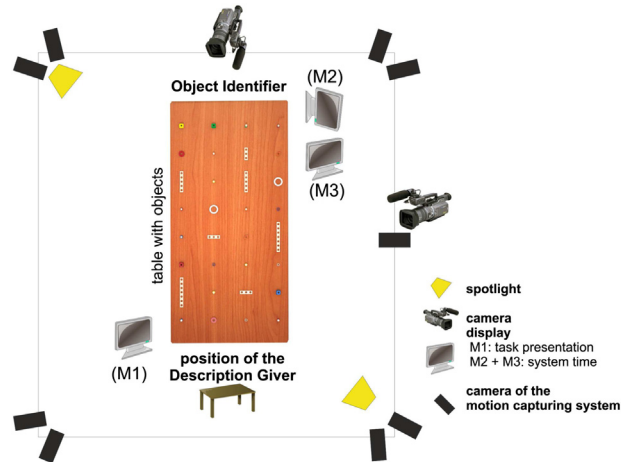


Fig. 4. Physical setup of the setting within the interaction space of the motion capturing system.

The feedback is restricted to ‘Ja’ (yes) in the successful case (accept) and to ‘Nein’ (no) in the unsuccessful case (denial). In both cases the current identification game exchange terminates and the participants move on, starting with the DG’s selecting the next object from his display. These steps of identification games constitute the level of *moves* used in data annotation. That is, these steps define functional units that give rise to a respective grouping of communicative behavior.

4.2. Participants

In total, 22 object identification games with speech and gesture and 22 object identification games with gesture only were recorded. Since each object identification game involves referring to each of the 32 objects, there are altogether 1408 ($2 \times 22 \times 32$) demonstrations to be analyzed. Each of the 44 pairs of participants completed *either* a S-G-Trial *or* a G-Trial, but not both. Gender is nearly equally distributed over the 88 participants: 41 female and 47 male volunteers joined the experiments. Their age ranges from 20 to 34. All participants were selected to be right handed.

4.3. Annotation

The main focus of annotation is the demonstration act of the DG. According to the two sub-settings – speech plus gesture or gesture only (see above) – two kinds of observable behavior are considered, namely verbal utterances (speech) and hand and arm movements (gestures). With regard to the latter, we follow the rough characterization of Kendon (1997) and take gestures to include only those hand and arm movements “that are treated by coparticipants in interaction as part of what a person meant to say” (p. 109 *et seq.*). Thus, “communicatively innocent” movements like stroking ones hair or rubbing ones earlobe are excluded from the beginning.

For OI, only the success of his identification is reported. Annotation was done on several layers by means of a coding scheme and affiliated coding instructions.

4.3.1. Annotation scheme

The annotation layers are described in the following list. They are partitioned into a gesture, a speech and a move domain. Participation in a domain is indicated by a respective prefix in the layer name. Sets of predefined annotation values are listed in brackets. Of course, annotating speech is restricted to the S+G-Trials.

- gesture.phase [*preparation, stroke, retraction*]: We adopt the structuring of a gesture motion according to the triple established by McNeill (1992). The interval in which hand and arm have their kinetic peak is called *stroke* (Kendon, 1980). From the linguist’s point of view, the stroke phase is the semantically marked one. The movement of the hand out of a rest position into the stroke is the *preparation* phase. Bringing the hand back from the stroke into a rest position (or into another gesture) is the *retraction* phase.
- gesture.handedness [*left, right*]: Is the gesture performed with the right or with the left arm? For two-handed gestures both values are specified.
- speech.transcription: DG’s speech transcribed at the level of words.
- speech.number: The number of words used in DG’s move.

- **speech.quality** [*type, shape, color, function, position, proxy*]: “semantic categories” that are referred to in an utterance. *Type* is used for descriptions that pertain to the nominal classification of objects, such as “screw” or “bar”. *Shape, color* and *position* are used to annotate utterances that refer to the respective properties of objects (say, “long”, “red” or “to the right”, to give examples for each in the order of mentioning.). Likewise, *function* is used in case the purpose or manner of use is a theme of an object description (as in “thing that you can plug into something”). Finally, *proxy*, labels taxonomically unspecified nouns, NPs or determiners, like “Ding” (*thing*) or “das” (*that*) or “dies Teil” (*this thing*). Occasionally, the DG refers to an object anaphorically by picking up some semantic dimension it shares with a prior referent from the discourse record. The annotators resolve such bindings and code the actual as well as the anaphorically transported types. If the DG uses an unspecified thingness-word as well as a specified type-declaration in one and the same demonstration – as it can happen, e.g., with repairs – only the more specific one, viz. type, is recorded.
- **move.referent** [α]: unique name α of the object referred to. Each object in the experimental domain has a unique internal name used during analysis. This name is used to keep track of the referent in the respective identification game.
- **move.success** [*yes, α*]: If the OI could successfully identify the object, the annotation value is “ja” (*yes*). It is name α for an erroneously chosen object otherwise.

Only those gesture tokens enter into analysis which are purely deictic (showing, e.g., no iconic traits like pointing and encircling synchronously in order to refer to a disk). Furthermore, the success (or failure) of a dialog move had to depend on exactly one gesture. We implemented this two-step filter using the following annotation layers:

- **gesture.validity** [*yes, no*]: Is the gesture a purely deictic one?
- **move.validity** [*yes, no*]: Is the game’s gesture valid and does the gesture include exactly one stroke?

Annotation of the video data has been carried out making use of two software tools, Anvil and Praat. The audio tool Praat (<http://www.praat.org/>) was used for the transcription of spoken language; the video films were annotated with the multimedia annotation tool Anvil (<http://www.dfki.de/kipp/anvil/>).

4.3.2. Reliability

As a test procedure for the reliability of the annotation scheme and its instructions the interrater-agreement between three raters’ annotations of one video on the most versatile layers, namely **speech.quality** and **gesture.validity**, has been calculated. As agreement coefficient, we use AC1 (Gwet, 2001), which measures agreement normalized for chance. Like the more widespread kappa statistics (Cohen, 1960), AC1 calculates agreement K in terms of the proportion of times annotators actually agree ($P(A)$) and the proportion of times annotators are expected to agree by chance ($P(C)$) according to the following formula (cf. Carletta, 1996):

$$K = \frac{P(A) - P(C)}{1 - P(C)}. \quad (1)$$

Agreement coefficients differ with respect to the heuristics they employ in order to estimate $P(C)$ (cf. Stegmann and Lücking, 2005). The chance estimator of the kappa statistics, $P(C_k)$, is computed as the joint proportion of the coding propensities of each rater (where the index ‘+’ denotes the column- respectively row-wise marginal sum of k annotations from the underlying contingency table):

$$P(C_k) = \sum_{i=1}^k \frac{n_{i+}}{n} \cdot \frac{n_{+i}}{n}. \quad (2)$$

Since $P(C_k)$ depends directly upon the annotation value proportions, it may lead to the paradoxical result of “high agreement but low kappa” in case of unbalanced distributions of those proportions (Feinstein and Cicchetti, 1990). One motivation behind AC1 is to reduce the risk of paradoxical results by computing the chance estimator, $P(C_{AC1})$, with reference to the concept of *intra-observer variation* (Kjærsgaard-Andersen et al., 1988), according to formula (3). Note that the risk is still not abolished completely, however.

$$P(C_{AC1}) = \frac{1}{k-1} \sum_{i=1}^k \gamma_i (1 - \gamma_i), \text{ with} \quad (3)$$

$$\gamma_i = \frac{(n_{i+}/n) + (n_{+i}/n)}{2} \quad (4)$$

For a more detailed discussion of these and related issues see Stegmann and Lücking (2005).

With a value of AC1 = 0.9 for semantic categories and a value of AC1 = 0.85 for gesture classification, both ratings prove to be quite consistent: Concordance of ratings cannot be ascribed to chance on a risk level of $\alpha = 0.01$. Most of the other layers have been evaluated extensively in an earlier study (see Lücking and Stegmann, 2005).

5. Results

5.1. Testing the (Inter Sem) hypothesis

Following the (Inter Sem) hypothesis we have to assume that every pointing gesture singles out exactly one object within the context c for all interlocutors – the experimental pointing domain, in our case. Given that (Inter Sem) holds, we should expect that the Description Giver is able to use pointing gestures to refer exactly to a specific object within c and that at the same time the Object Identifier can easily identify the object, especially within the small domain used in our setting (see Fig. 3). However, this is not what we find and we demonstrate that taking the Object Identifier's perspective.

In the G-Trials (only gestures are allowed, see section 4.1) from the Object Identifier's perspective the pointing gestures fail in the distal area starting from row 4 onwards in about 28.3% of the cases. This can clearly be seen in Fig. 5 showing the number of failed identifications per row. In the first three rows, all pointing gestures have been identified as referring to exactly the correct object. In row 4 we find 4 failed identifications and the count increases from there on. Looking at the statistics alone, the drop of identification errors in row 8 is surprising. Reviewing the video recordings, we realized that a specific border-of-the-domain behavior can be found for pointings to this row: the Description Giver exaggerates the pointing gesture by pointing clearly beyond row 8. Using this strategy, dubbed “gestural hyperbole” in Kranstedt et al. (2006c), the Object Identifier only has to discriminate between the 4 objects in the row, and does not have to consider the neighboring rows. This seems to be a quite successful strategy, as the results attest. This result also shows that an empirical model of referring by pointing even in concrete domains has to make provision for allowing for exceptions, marked instances or special context conditions. Bearing this *proviso* in mind, we are careful in spelling out our empirically informed pointing rule (INF) below by making it defeasible and acknowledge overriding constraints or principles.

In the S+G-Trials (speech and gestures are allowed) we find a rate of 99.8% of successful identifications. Since the production and interpretation of both gesture and speech contribute to the success of the identification, we cannot draw a direct conclusion for the role of the gesture from this aggregated result. In the S+G-Trial the pure number of failures is not a valid indicator for the performance of the pointing gesture. However, looking more closely at the annotated data, we find a pattern similar to that in Fig. 5, if we count the mean number of words used during a demonstration for each row, as shown in Fig. 6. During a demonstration within the first three rows, three words are used on average. The mean number of words then increases to 6 words for the rows 6, 7, and 8. Thus, the Description Giver seems to compensate for the increasing fuzziness of his pointing gesture at distance by producing a more elaborated verbal expression. It could still be the case that the Description Giver believes his distal pointings to perfectly refer to the objects, but he is at least aware of the perspective distortions the Object Identifier has to cope with and employs an audience design in his utterance. This strategy is quite effective, as the success rate of 99.8% suggests. Strategies of this kind have in common that pointing gestures are altered in order to cope with the loss in expressiveness of verbal expressions either due to the lack of speech (G-Trial) or due to the incomplete meaning of deictic expressions (S+G-Trial). A more in-depth analysis of the behavior of the Description Giver of Pfeiffer (2012) showed more of such strategies in an interaction of the modalities speech and gesture.

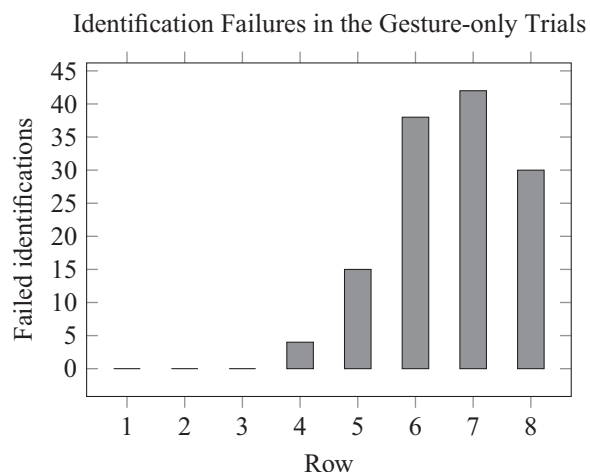


Fig. 5. Object Identifier's perspective: The number of identifications the Object Identifier failed per row increases from row 4 onward in the G-Trials (see annotation *move.success*). For the Object Identifier, pointing gestures fail to refer to a single object in the distal area (measured from the position of the Description Giver). In the S+G-Trials we observed only one identification failure in 704 demonstrations.

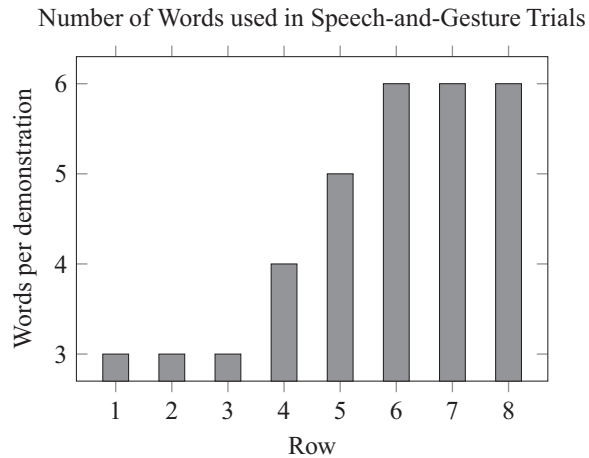


Fig. 6. Description Giver's perspective: The number of words the Description Giver uses per row increases from row 4 onward in the S+G-Trials. The Description Giver uses speech to compensate for the fuzziness of the pointing gesture in the distal area (Lücking et al., 2004).

We have thus provided evidence that, even within small distances, pointing fails to successfully refer to a single object for the recipients in the general case. Remember that the distal area is already starting at a distance as short as 68 cm in our trials. Taking the data from both trials into account, we have to drop the (Inter Sem) hypothesis. This is in accordance with Butterworth and Itakura's findings concerning pointing vectors (Butterworth and Itakura, 2000).

5.2. Testing the (Intra Sem) hypothesis

So far we have examined the overall success of a demonstration, defined by a correct identification of the demonstrated object by the Object Identifier. It could still be, e.g., that it is only the Object Identifier, for which the (Sem) hypothesis, and thus (Inter Sem), fails. For the Description Giver the (Intra Sem) hypothesis might still hold. This is what we intuitively expect, because the Description Giver is the one pointing to the object.

This motivated us to have a closer look at the pointing gesture itself. Testing (Intra Sem) on the pointing gesture means that an operationalization of the hypothesis in terms of the spatio-temporal features of the pointing gesture and the context is needed.

We start with a rigorous operationalization, trying to capture the rigidity of the (Sem) hypothesis. In our setting, the Description Givers used the prototypical extended index finger to point to the objects. We therefore say a pointing gesture refers to the first object that is hit by a ray shooting from the tip of the index finger (see Fig. 7).

Operationalization 1. Rigorous Operationalization of the Sem Hypothesis: A pointing gesture refers to the first object which is hit by a pointing ray extending from the index finger.

Based on the measurements of the tracking system we are able to reconstruct the position and orientation of the index finger during each stroke. In addition, we also have exact knowledge of the positions of the objects on the table. Thus we can, for each demonstration, project the ray from the index finger at the time of the stroke and geometrically test, whether the ray hits any of the objects.

Before we do this, we have to stipulate at least two parameters for the ray: its anchoring and its orientation. Since the subjects use prototypical index finger pointing, the anchoring is located on the tip of the index finger. Yet there are several

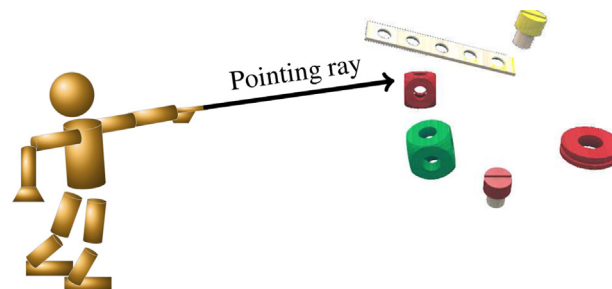


Fig. 7. Assuming the rigorous operationalization, the pointing gesture refers to the smaller cube below the 5-hole bar.

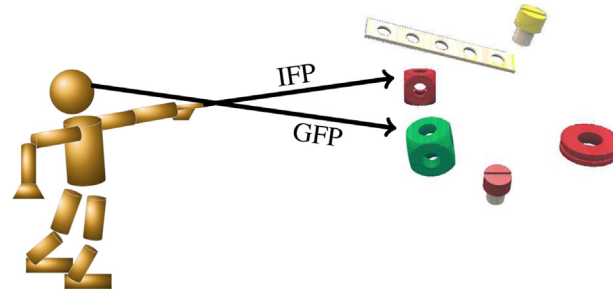


Fig. 8. The figure shows the two most prominent options for determining the direction of the ray considered in this paper. In the index-finger-pointing (IFP) approach the ray extends along the orientation of the index finger. In the gaze-finger-pointing (GFP) approach it prolongs the direction of gaze aiming at the target over the tip of the finger. As is shown in the picture, adopting IFP or GFP might result in different assumptions regarding the object hit by the pointing ray: the red cube if IFP is adopted, the green cube if GFP is adopted. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Table 1

The table subsumes the errors made according to the IFP or GFP model in terms of (a) differences in x, (b) differences in y, (c) differences in angles and (d) differences in minimal distance each between the IFP- or GFP-ray and the object. For each difference the mean (expressing accuracy) and the standard deviation (expressing precision) are given. IFP values and GFP values are separated by a solidus: IFP/GFP.

Row	1	2	3	4	5	6	7	8
<i>x (cm), row</i>								
Mean diff.	-3.1/-2.4	-5.5/-0.8	-4.1/1.7	-4.8/11.8	-0.9/19.7	20.3/33.7	36/54.2	82.9/105.8
Standard dev.	±7.1/±4.7	±3.5/±4.8	±4.8/±5.6	±6.7/±24.8	±11.7/±24.8	±54.1/±51.8	±87.4/±43.5	±99.3/±94.7
<i>y (cm), col.</i>								
Mean diff.	4/1	4.3/1.3	4.8/1.9	5.1/-0.6	6.9/-3.8	9.3/-6.3	12.9/-16.2	9.9/-38.7
Standard dev.	±4.8/±6	±3.5/±5.8	±2.7/±7.6	±3.7/±16.7	±6/±17.6	±20.8/±19.6	±28.9/±21.9	±22.5/±49.2
<i>Angle (°)</i>								
Mean diff.	33.8/32.1	30.9/22.8	24.6/16.3	17.7/16.7	12.5/13.4	13.5/12.5	10.1/12	10.3/12
Standard dev.	±12.8/±16.9	±10.6/±13.9	±8.9/±14.5	±8.9/±14	±6.8/±11.9	±6.2/±12.1	±4.7/±7.1	±4.8/±7.3
<i>Ortho (cm)</i>								
Mean diff.	6.1/5.9	7/5.4	6.9/5	7.9/8.3	8.9/9.3	12.3/10.4	12.8/14.9	17.6/19.9
Standard dev.	±4/±4	±3.9/±4.8	±6.3/±7.1	±7.4/±9.1	±8.4/±8.8	±9.5/±9.7	±6/±8.6	±9/±8.4

possible ways to determine the direction of the pointing ray. It could be determined solely by the orientation of the index finger, we call this style of pointing *index-finger-pointing* (IFP). The IFP can be interpreted as a body-extension model of the pointing ray. Alternatively, the direction can be determined by the direction of gaze, aiming at the target over the tip of the finger (we call this *gaze-finger-pointing*, GFP). The GFP model includes the perspective distortions the Description Giver experiences. Hence we here integrate an assumption of David Kaplan's (Kaplan, 1989a:525–526). Thus, if the Description Giver visually checks whether he is accurately pointing to an object or not, this is the most plausible model for a pointing ray. It could, however, be anything in between, e.g., depending on the laterality of the object pointed to. In the following, we will consider both, IFP and GFP, as the most prominent exponents of the set of possibilities (see Fig. 8).⁹

We get distributions of points around the target objects showing precision and accuracy of the pointing gestures (see Fig. 9) by applying these pointing models to the recorded data and intersecting the generated rays with the surface of the table.

Looking at Fig. 9 and Fig. 10, we find that pointing is fuzzy. In most of the demonstrations the projected ray completely fails to hit the target object. Looking at object (4/1), we even find that the object pointed to is actually lying outside the 75% boundary of all pointing strokes referring to it. A quantitative description of the precision and accuracy of the pointing gestures for both IFP and GFP is shown in Table 1. The fuzziness can be observed both in the distal (rows 4 to 8) as well as in the proximal area (rows 1 to 3). The precision of the ray, expressed by the standard deviation, decreases with the distance of the object from the Description Giver and thus the areas covered by the bagplots are getting larger. The same

⁹ In his book "Understanding Multimodal Deixis with Gaze and Gesture in Conversational Interfaces" (Pfeiffer, 2011), Pfeiffer presents results based on a data-driven modeling of pointing origin and pointing direction using the data elicited in the presented study. In his work, he substantiates the claim for GFP and, moreover, shows that a context-specific dominant eye defines the direction of pointing together with the tip of the index finger. However, even with this more precise model of pointing, the claims made above still hold.

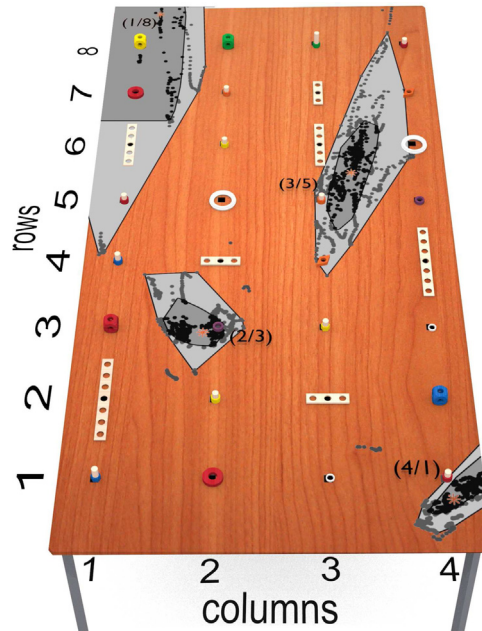


Fig. 9. The distribution of the context objects on the table is overlaid with bagplots visualizing the precision and accuracy of the pointing gestures found for four selected objects (indicated by the pair of coordinates in (column/row) notation as seen from the DG's perspective) in the G-Trials. Thus, the object in position (3/5) is the orange slit bolt. Within each bagplot the dots mark the intersections of a pointing ray with the surface of the table. The star indicates the mean position. The area with the darker shading includes 50 percent and the area with the lighter shading 75 percent of the points. In this setting, the Description Giver was sitting on the lower side, the Object Identifier was standing on the upper one. It can clearly be seen that most of the rays fail to hit the target object. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

can be observed for the accuracy: the distance of the mean (marked by the asterisk in Fig. 9) over all demonstrations to a specific object also increases with the distance of the demonstrated objects to the Description Giver. A pointing's deviation is assessed in terms of two measures: *angular deviation* and *orthogonal deviation*. Angular deviation is the angle between the stipulated IFP/GFP-ray and an ideal straight connection of the index finger and the center of the object. Orthogonal deviation is the distance between the position of the object in question and its orthogonal projection onto the IFP/GFP-ray. Note that angular deviation is a distance-independent measure while orthogonal deviation is not (Fig. 10).

This analysis provides us with an explanation for the effect represented in Figs. 5 and 6: in the proximal area accuracy and precision are reasonably smaller than the distances between the objects. The Object Identifier can easily assess the correct demonstrated object, and thus both measures, the number of fails (Fig. 5) and the number of words (Fig. 6), show no variation. With increasing distance accuracy and precision are getting lower than necessary to differentiate between neighboring objects and thus the indices (Fig. 5 and 6) increase.

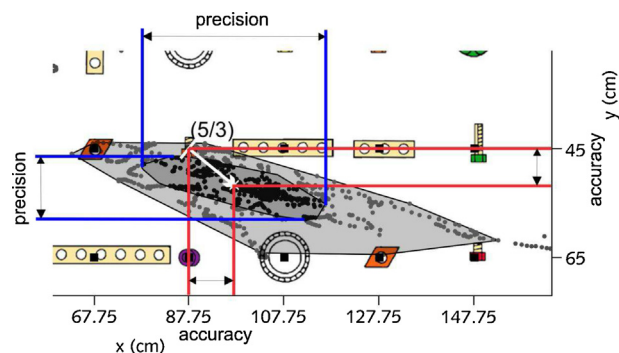


Fig. 10. This graphic provides a detail from Fig. 9. It exemplifies the concepts of accuracy (difference between the mean of the data and the ideal pointing ray) and precision (spreading of the data around the mean). In the tables these values are expressed in terms of the mean difference and the standard deviation.

If we take pointing rays under IFP or GFP as *explicans* for the *explicandum* “demonstrated by *p*” (Davidson) or “demonstration δ ” (Kaplan) we see that there are too many fails in particular in the distal area in order to serve as an appropriate *explicans*. Any *explicans* provided has to be sensitive to the position of the pointer, her gauging apparatus (IFP/GFP) and the density of the domain pointed into. In order to secure that a demonstration succeeds in a dense domain, the pointer may have to shift her position or lower her torso towards the area pointed into. The IFP or GFP machinery has to change accordingly with the effect that the emitted ray can hit the target. Nothing of this dynamics is tied up with Davidson’s *demonstrated* or Kaplan’s *demonstratum*.

Assembling the results for the proximal and the distal area, we have to reject both versions of the (Sem) hypotheses tested against the rigorous operationalization. The rejection depends primarily on the strict ray model of pointing (recall Fig. 7 for an illustration).

5.3. Coping with the low precision of pointing

In the last section, we took the position that we need a rigorous operationalization of the (Sem) hypothesis. Reconsidering this in the context of the results shown in the bagplots (Fig. 9), a more relaxed operationalization could be suitable to compensate for the low precision of pointing, while still allowing us to sustain the (Intra Sem) hypothesis.

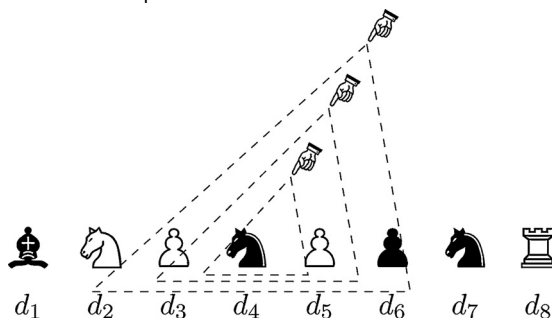
We are thus looking for a model for the pointing extension which is more suitable to grasp the precision of pointing than the pointing ray does. Looking closer at the distributions of the intersections between the pointing ray and the surface of the table (Fig. 11), one can see that they seem to approximate ellipses. Earlier studies suggested a cone as a potential model of the pointing extension (Lücking et al., 2004).¹⁰ Mathematically, when intersecting a cone with a surface, the result is always an ellipsis (including circles as the result of the special case of strict orthogonal intersections). Following this idea, we fitted ellipses onto the pointing data for every demonstrated object, with visually plausible results (Fig. 11). This motivates us to formulate a relaxed operationalization of the (Sem) hypothesis using a pointing cone to model the low precision of pointing, as illustrated in Fig. 12.

Operationalization 2. Relaxed Operationalization of the Sem Hypothesis: A pointing gesture refers to the object which is hit by a pointing cone extending from the index finger.

To investigate the pointing cone variant of (SEM), we calculated the optimal opening angles for the IFP and GFP variants of the pointing cones on a per row basis. The performance of these cones was measured by computing the percentage of successfully singled out objects based on the recorded data (see Table 2). Overall, the results show that with a fixed opening angle the performance of this method is below 50% in our widely spaced domain and it can be expected to be even lower in more crowded domains. For the proximal objects quite large opening angles are needed to include the objects referred to at all, while for the distal objects more restricted opening angles are needed to exclude false positives. Targeting at an overall optimum for our domain, quite large opening angles of 71° (IFP) respectively 61° (GFP) are needed to achieve a maximum rate of success of 38.5% (IFP), respectively 48.1% (GFP), of successfully identified demonstrations to objects. These rates of success are too low to provide a feasible foundation even for a weaker (Intra Sem).

Hence we are led to the conclusion that pointing is not a semantic referring device. A Wittgenstein–Davidson–Kaplan inspired view on demonstration by pointing seems to be out.

¹⁰ Work with pointing cones in order to model the demonstrata of pointing acts has a fairly long tradition in Bielefeld gesture research (see Kranstedt et al., 2006b,c for further references): the intersection between the cone generated with a pointing instrument and a surface is used to estimate the precision of pointing. Roughly, precision decreases with the distance from the pointing source. An alternative conceptualization of the precision issue we are aware of is van der Sluis and Kraemer (2007). They use a “Flashlight Model” to account for various types of pointing gestures of different precision. In order to capture the difference between these two styles of modeling compare their Fig. 5 – for sake of convenience reproduced below:



– with Figs. 12 and 13 and the measurement data in Table 3 in this paper.

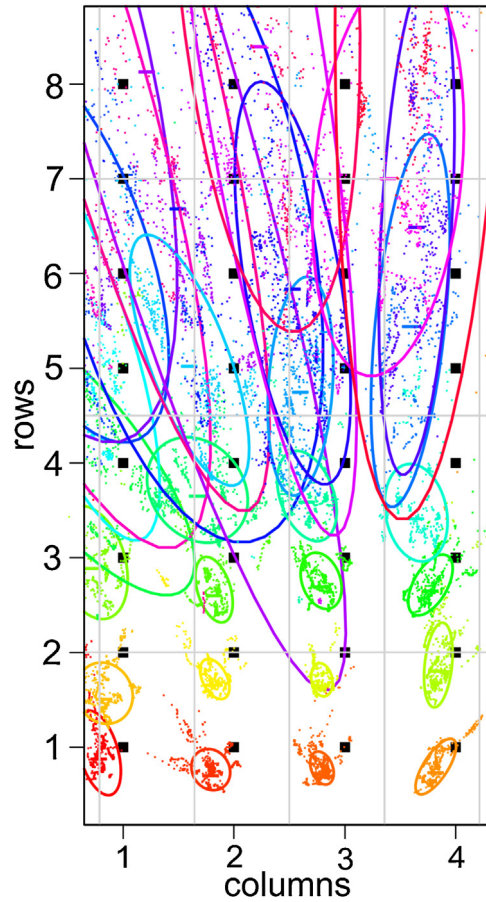


Fig. 11. The distributions of the intersections between the pointing ray and the surface of the table, here shown for the G-Trials, can be approximated by ellipses. In the figure, ellipses are representing normal-probability contours of level 0.75. The rectangular markers represent the centers of the ellipses.

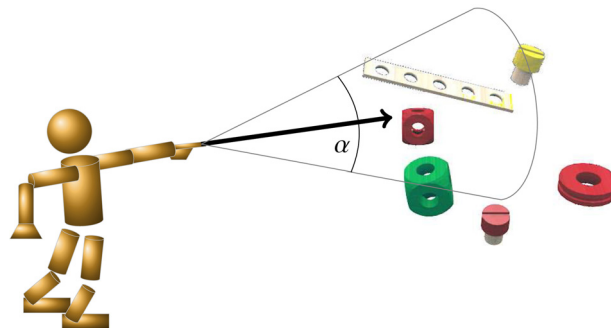


Fig. 12. A pointing cone can be used to model the precision of the pointing which decreases with the distance. Based on this model, pointing refers to the object hit (or intersected) by the cone.

5.4. Testing the pragmatic hypotheses

In the last section we have seen that pointing is fuzzy and we thus could not manage to maintain the semantic hypothesis, even when we relaxed our operationalization and used the less restrictive pointing cone instead of a pointing ray. For the semantics interface we would need a test for reference providing a definite single object for every demonstration. In pragmatics this is different. Here we can use inference to choose between objects from a set of possible referents, selecting the one most likely intended, which led us to formulate the two pragmatic hypotheses in section 3.

Table 2

The optimal opening angles α for the IFP and GFP variants of pointing cones per row and their performance in terms of percentage of successfully singled out objects per row. The GFP variant shows a slightly better performance than the IFP variant.

Row	1	2	3	4	5	6	7	8	All
<i>Index-Finger-Pointing</i>									
Angle α (°)	84	80	71	60	36	24	14	10	71
Performance (%)	70.27	61.84	71.43	53.95	43.84	31.15	23.26	7.14	38.54
<i>Gaze-Finger-Pointing</i>									
Angle α (°)	86	68	69	38	24	25	17	10	61
Performance (%)	68.92	75	81.82	65.79	57.53	42.62	23.26	14.29	48.12

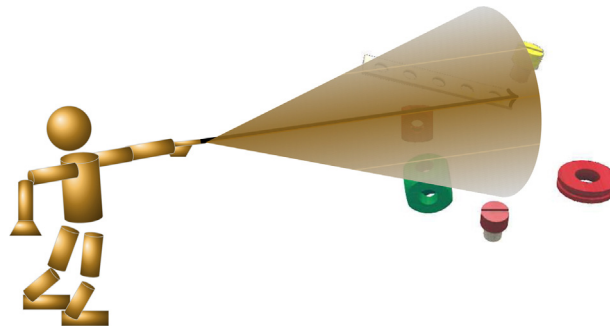


Fig. 13. The pragmatic-based operationalization model combines a pointing cone with a simple weighting heuristics (shading), which prefers objects closer to the axis of the pointing cone (which can be compared to the pointing ray).

The argument for the (Weak Prag) hypothesis is straightforward: we observe 83.9% of successfully identified pointings to objects in the G-Trials. Thus we have to conclude not only that the attention of the Object Identifier is actually drawn to the target object by pointing, it also proves to be a quite successful strategy.

To examine the (Strong Prag) hypothesis we will only use motion capturing data. We modify our relaxed operationalization for testing the (Sem) hypotheses and allow several objects to lie within the pointing cone, as long as the intended target object can be singled out from the set of objects delimited by the pointing cone in a subsequent inference.

Operationalization 3. Operationalization of the Strong Prag Hypothesis: A pointing gesture refers to the object identified by an appropriate inference from the set of objects covered by a pointing cone extending from the index finger.

For the inference process, cf. (INF) in section 6, we start using a simple heuristics ranking the delimited objects according to their distance from the central axis of the pointing cone. We use an angular measure instead of Euclidean distance because the angular measure is independent of the perceived distance. Otherwise, objects farther away would have a larger Euclidean distance than proximal objects, although they would appear to be closer to the central axis. In Fig. 13 the new pragmatic-based operationalization model for the extension of pointing is depicted. The distance dependent heuristics is visualized as a shaded area, emphasizing the objects nearer to the central axis of the cone.

Again, we computed the optimal opening angles and the performance in terms of percentage of successful identifications per row (see Table 3) using computer simulations driven by the tracking data recorded in the study.

Table 3

The pragmatic-based operationalization model combining a pointing cone with a weighting heuristics improves performance. The table shows the optimal opening angles α for the IFP and GFP variants of pointing cones and their performance in terms of percentage of successfully singled out objects per row. Now the IFP variant shows a slightly better performance than the GFP variant.

Row	1	2	3	4	5	6	7	8	All
<i>Index-Finger-Pointing</i>									
Angle α (°)	120	109	99	109	72	44	38	31	120
Performance (%)	98.65	100	94.81	98.68	97.26	91.8	86.05	52.38	96.04
<i>Gaze-Finger-Pointing</i>									
Angle α (°)	143	124	94	89	75	50	41	26	143
Performance (%)	98.65	100	93.51	93.42	94.52	90.16	67.44	69.05	92.71

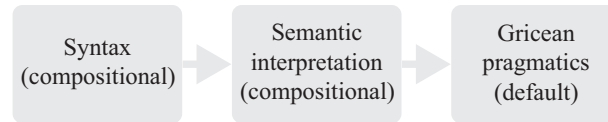


Fig. 14. The classical input–output model of the semantics–pragmatics interface.

Using the straight-forward weighting heuristics, the pragmatic inferences succeeded in 96% of the cases of the full domain when using index-finger-pointing and in 93% of the cases when using gaze-finger-pointing. The large optimal opening angles suggest that these excellent results are mainly due to the weighting heuristics and not due to a clear-cut cone intersection.

The model performs even better than the human Object Identifiers in our study. This can be attributed to the semi-omnipotent knowledge of the computational approach: based on the motion capturing data the exact position of the relevant body parts during the pointing gesture can be used to compute the extension, while the human Object Identifier only has a perspectively distorted view.

At the bottom line, a pragmatic inference process based on an elementary weighting heuristics provides sufficient means to disambiguate pointing in our setting. We cannot tell from these results whether either the human Object Identifier or the Description Giver uses heuristics based on an angular distance measure. However, we can tell that the pragmatic interpretation of a pointing gesture does work, as opposed to approaches purely based on a strict semantic interpretation.

Hence, we take this as evidence that (Strong Prag) holds.

The pragmatic rendering of demonstrations according to (Strong Prag) intuitively diverge sharply from the deterministic idea of direct reference. Notwithstanding this impression, an inferential treatment of reference poses indeed deep problems in particular for a Kaplanian approach as well as for a direct reference view in general. The reasons are as follows. We can think of (INF) in such a way that it becomes operative in situations where there are several (i.e., more than just one) objects in the region covered by the pointing cone. Otherwise – that is, if there is just a single possible demonstratum – (INF) becomes trivial and can be skipped. In this case, but only in this case, demonstrating by means of a pointing cone can be considered to be a relaxed variant of a directly referential beam projecting model (cf. the respective discussion in the preceding section). However, as has been revealed in our experimental study, pointing cones concern a collection of objects. In such situations, a strict or relaxed direct reference rule fails. The failing of some semantic rule triggers the application of an inference in the first place, in this case, a perceptual inference from idealized beams to objects. “Inferential reference”, as we may coin it, therefore contradicts Kaplanian principles driving demonstration in two ways: firstly, if no demonstratum is singled out immediately, a demonstration δ simply counts as failed. This follows from the semantic rule associated with complete demonstratives (cf. above or Kaplan, 1989a:529), which has an *if*-clause capturing non-demonstrating demonstrations that applies in this case.¹¹ Secondly, employing an inferential rule in addition to demonstration δ violates Kaplan’s first “obvious principle”, which states that “the referent of a demonstrative depends on the associated demonstration” (Kaplan, 1989a:492). The restricting condition necessary for this second argument, *videlicet* that the referent depends *exclusively* on the demonstration, is given by Kaplan some pages later, where he claims that a demonstration in a context *determines* the demonstratum (Kaplan, 1989a:526). Taken together, the Kaplanian standard view on demonstratives states that the demonstration act is both the necessary and sufficient condition for identifying the demonstratum. This clear and decided view has to be given up if one concedes that inferences are an indispensable auxiliary: the whole idea of demonstratively singling out objects by direct reference seems to rest on a myth. It is evident that both, a semantic and a pragmatic explanation for pointing have to account for demonstrative reference fails. But they do so in entirely different ways and with different effects for the semantics–pragmatics interface. Kaplan’s *demonstration* works like an assignment or an interpretation function and is thus wholly situated in semantics proper, since determined by the model. This is in accordance with the classical view concerning semantics which we presented in Fig. 14 when bringing up foundational issues in section 6. What we must suggest, however, given the statistical variance results we have, is that an inference is at work to yield a referent. This in turn yields an entirely different model of a semantics–pragmatics interface, namely, pragmatic information computed as a precondition for compositional semantics. We discuss some consequences of adhering to (INF) rather than to a Kaplanian model in the subsequent section 6.

6. Foundational issues: pointing experiments and the set-up of a semantics–pragmatics interface

In this section, we will investigate the impact of the empirical findings presented in sections 2–5 on the role of the semantics–pragmatics interface within descriptions of language. Here is a brief summary of the state arrived at so far:

¹¹ Additionally, note that Kaplan explicitly allows for failed demonstrations (Kaplan, 1989a:525).

what we have found out is that pointing does not directly single out an object by some sort of pointing beam. As a consequence, we cannot liken pointings to interpretation functions for constants which automatically provide a value. As a next step, we will investigate in more detail how pointings will fare with respect to the suggestions of Davidson and Kaplan concerning complex demonstratives, cf. section 3.

Obviously, there is no demonstrated object in the sense of Davidson and no *demonstratum* in the sense of Kaplan wrt. the pointing-ray model – see sections 4 and 5.

This leads us to the conclusion that Davidsonian truth conditions are not meant for natural languages but only for formal languages. This, however, would be adverse to the Davidsonian program (cf. Davidson, 1967). Anyway, perhaps we can do better using Kaplan's proposal, since the wording of Hypothesis 3 seems to open up an additional possibility: he relates the demonstration δ going into the character of a complex demonstrative to some context c which we are free to restrict.

Since satisfaction problems seem to be mainly tied to the distal area, one idea would be to restrict Kaplan's satisfaction condition to the proximal area of the Description Giver's pointing domain, say to the first two rows. Then we can generalize the notion of a "proximal c " for Kaplanian theories, narrowing down his context c to something like "the proximal action space c of the speaker". This amounts to claiming that Kaplan contexts always have to be proximal contexts. Now clearly, a general notion of proximity would be hard to define but we do not need to do that here. This restricted notion of a context would agree with the interpretation of Kaplan contexts as narrow contexts which one can find in the literature (cf. Recanati, 2004b). However, if we look at Fig. 3 (4/1) we see that our data does not permit this elegant move. Looking at the first three rows in Fig. 3, we notice that there does not seem to be an easily accessible way to define restricted Kaplan contexts c . As a last resort, we could try to establish restricted contexts c as those which satisfy objects demonstrated but that would maneuver us into a vicious circle.

So, semantic modeling seems to fail with respect to our data and the pragmatic hypothesis already introduced in section 5.4 is what we must try to put to work. We take the relevant interpretation of (INF) (defined below) as some sort of default. It has a high success rate as Table 3 shows. So, why worry at all, given the flexible default notion? Well, because now the worries come from a different side, and, so, finally, we arrive at the methodological question hinted at in the introduction, which we will now set out to discuss.

Let us start with some Description Giver's request like "Grasp the red slit-screw", accompanied by one of the pointings (i. e. pointing act tokens) as shown in (4/1), Fig. 3. In the following, we by and large assume a compositional treatment of the semantics of speech acts as proposed, e.g. in several writings of Vanderveken (cf. Vanderveken, 1990). Observe that – neglecting for a moment the issue of complex demonstratives and turning to the wording of the verbal description – if we consider only the definite description, it also cannot be satisfied, because there is a second red slit-screw in position (1/5). So, given the classical semantics and its Davidsonian or Kaplanian variants, we do not arrive at an object referred to, since reference fails. Motivated by successful pointing actions in our setting, however, how can we accommodate that in the classical schema of the semantics–pragmatics interface? The classical interface can be pictured in the input–output-model from Fig. 14.

Depending on whether we opt for a presuppositional or a non-presuppositional (perhaps Russellian) approach for the complex demonstrative in the request, we can either say that the request cannot possibly be satisfied due to presupposition failure or that it cannot be satisfied on purely semantic, i.e. referential grounds. Using an analogy with assertives, we conclude that, if we have a presupposition failure, no felicitous request can be expressed. Again, the non-presuppositional argument is that the complex demonstrative will lack a referent. Since Gricean approaches are essentially bi-valued, we follow up the second line of argument here and assume that the request cannot be satisfied. We might then try to work on the assumption that Grice's Quality maxim is violated. As a consequence, we must establish a sound argument to arrive at a felicitous implicature which then could provide us an *implicatum* with a working request. It is not clear how this should be brought about in detail or whether it could be put to work at all. What are the problems with such a procedure? Two immediately come to mind: First, on a Gricean account we must be clear about the level of 'what is said', which we are not, because of the reference failure. Hence, we cannot apply Grice's working schema (Grice, 1975; Harnish, 1991) for implicatures. Second, once we have started Gricean routines, we cannot move back into semantics again: Gricean routines do not loop back. The third problem is perhaps a bit more difficult to see due to various sloppy readings of Grice around: 'What is said' is essentially propositional and our problems arise at the sub-propositional level.

As obvious from Fig. 14 above and the VP in the example "Grasp the red slit screw", we want to have a suitable reference for "the red slit screw", in order to get a denotation for the whole VP. However, as the empirical results in section 5.3 show, the only way to get an object is to use a default¹² inference. The pragmatic operationalized model derived in section 5.4 leads to the following strictly formulated (INF) hypothesis:

¹² Recall from the discussion of Fig. 5 from section 5.1 that we already encountered a defeasible strategy overriding (INF), namely "gestural hyperbole" (Kranstedt et al., 2006c). For profiling further pointing strategies see Pfeiffer (2012). For these reasons (INF) is formulated as a default rule here, allowing for incorporating such "strategic pointing behavior".

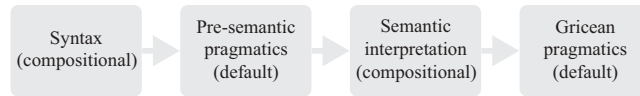


Fig. 15. The input–output-model of the updated semantics pragmatics interface.

Hypothesis 6. (INF): We say that an object is referred to by pointing only if

- (a) the object is intersected by the pointing cone and
- (b) the distance of this object from the central axis of the cone is less than any other object’s distance within this cone.

It is noteworthy that we do not claim that description giver and object identifier can spot the central axis of the pointing cone. Rather, the INF rule is a pragmatic rule which has the best fit for our empirical data, the pointings to objects on the table, which we can provide. Similarly, the pointing vector does not approximate a psychological predicate. We haven’t elicited subjects’ reports about their gauges of pointing vectors but rely on measurable ray-table intersections. It may be the case that the description giver misconceives her own produced pointing gestures as directly pointing towards the target and the pointing cone describes the variance of this misconception. In this sense, the cone can be seen as representing the distribution of the incapacity of speakers to bring about precise pointing (precise in terms of vector extrapolation) in the first place. Following this understanding, the cone is an abstract but data-driven model of pointing behaviors. As such, it can be reified to give rise to a theoretical entity, as in (INF). Hence, the pointing ray can be conceived of as an operationalization of, respectively, Wittgenstein’s notion of pointing, Davidson’s meta-language demonstrated and Kaplan’s demonstration δ . This operationalization furthermore shows that even if interlocutors would be able to precisely render pointing rays, these rays won’t usually hit their targets and therefore are also not sufficient for providing a cognitive account for demonstrative reference.

According to INF, now, the object demonstrated is the one nearest to the axis of the pointing cone as Fig. 13 shows. Conditions (a) and (b) are considered necessary conditions.

The only way out of our dilemma seems to be administering (INF) wherever it is needed to produce a suitable value. Then, however, Fig. 14 above cannot be upheld. “Application wherever needed” provides us with a non-standard picture of the semantics–pragmatics interface, namely the one displayed in Fig. 15.

Pre-semantic pragmatics might give us the object demanded. In Fig. 3 (4/1) this will be the red slit bolt on the fringe of the weaker base. Given this object, the semantics of the VP and the request can be computed compositionally. So, we end up with compositional procedures after all. Observe that since we get a denotation for the complex demonstratum we also circumvent the problem of the two slit-screws mentioned above. Obviously, we work with two different semantics–pragmatics interfaces now: one of them is pre-semantic or perhaps “woven” into semantics and serves as a precondition to compositional semantic procedures. For simplicity, we will call these modes of operation “pre-semantic pragmatics”. Pre-semantic pragmatics will consist of a set of default rules like (INF) – observe that these are not Gricean. The second pragmatics slot is in a post-compositional semantics position, since we still could need Gricean pragmatics in order to arrive at the request via implicature. This would, e.g. be the case, if we had an indirect speech act like “Now, you grasp the red slit bolt” being an assertion.

Recanati (2004a) provides the most detailed discussion of the semantics–pragmatics interface existing to date from the philosopher’s perspective. Before we present one of his diagrams showing the relation between semantics and pragmatics, a brief quotation illustrating his findings is in order:

To sum up, either semantic interpretation delivers something gappy, and pragmatic interpretation must fill the gaps until we reach a complete proposition. *Or we run semantic interpretation only after we have used pragmatic interpretation to pre-determine the values of semantically underdeterminate expressions, which values we artificially feed into the narrow context.* Either way, semantic interpretation by itself is powerless to determine what is said, when the sentence contains a semantically underdeterminate expression. (Recanati, 2004a:58, emphasis added)

The passage written in italics corresponds to the strategy we arrived at exclusively based on the pointing experiments and the resulting measurements. What this amounts to can be explained with respect to Recanati’s diagram (Figure 4.1, p. 52 in his book, see Fig. 16, left) and our minimal revision of it (see Fig. 16, right).¹³

Recanati’s *Sentence meaning* captures the semantic content of the sentence in the strict formal sense. *Saturation* is a primary pragmatic process that cares for the mandatory, minimal and linguistically triggered interpretation of pronouns,

¹³ See footnote 4 on page 6 for a brief note on (Kaplanian) narrow contexts.

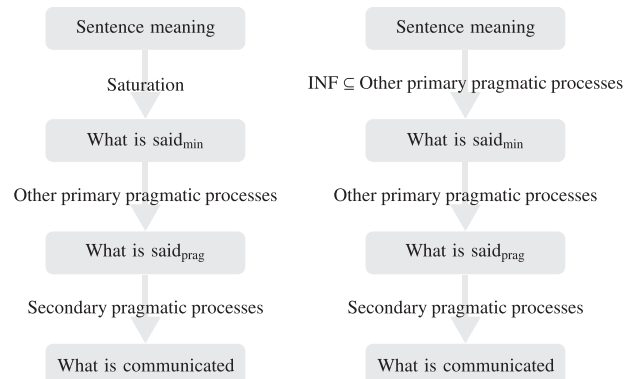


Fig. 16. Recanati's diagram (left) showing what he calls the *Syncretic View* concerning the relation between semantics and pragmatics. This view corresponds to the standard input–output schema of interpretation (see Fig. 14). The right diagram shows our extension in which we added an inference process in the transition from Sentence meaning to *What is said_{min}*.

indexicals, definite descriptions etc. via an interpretation function and is provided by the narrow context. Then we arrive at *What is said_{min}*. Albeit this is already a complete proposition, it will, as a rule, need *Other*, “higher-level”, optional primary pragmatic processes, in order to enrich the logical form and produce *What is said_{prag}*. On this pragmatic content Gricean mechanisms of the inferential sort such as rules for conversational implicatures, called *Secondary pragmatic processes* here, can operate and we get *What is communicated*.

In light of our experiments, this syncretic view (Recanati's wording) has to be slightly modified. (INF) is not simply a contextual interpretation function since it involves a “higher-level” default inference. Hence, it cannot be seen as an instance of a saturation process. Instead, referring into the perceptible environment shows that *Other primary pragmatic processes* already operate on the most basic interpretation stage in order to yield a proposition, i.e. *What is said_{min}*.¹⁴

7. Looking at other paradigms

In this section we will briefly look at how pointing can be captured within alternative frameworks that deviate from the classical theory by building in semantic indeterminacy right from the start. We treat three lines of research in this context, namely the language-as-situated-activity program (Barwise, 1989), Relevance Theory (RT; Sperber and Wilson, 1986), and the explanatory reference model of Roberts (1993). The conception of truth underlying these three semantic accounts can be traced back to Austin (1950). His notion of “true” is opposite to that of Tarski (1935, 1944) which is fundamental for the classical theory. Austin pairs predication (“descriptive conventions”) with an additional anchoring mechanism (“demonstrative conventions”). In Austin's own words (footnotes omitted):

[...] there must be two sets of conventions:

Descriptive conventions correlating the words (= sentences) with the *types* of situation, thing, event, &c., to be found in the world.

Demonstrative conventions correlating the words (= statements) with the *historic* situations, &c., to be found in the world.

A statement is said to be true when the historic state of affairs to which it is correlated by the demonstrative conventions (the one to which it ‘refers’) is of a type with which the sentence used in making it is correlated by the descriptive conventions.

Apparently, an Austinian semantic theory entails a semantics/pragmatics distinction that diverges from the standard input–output model (see Fig. 14). The diverging set-up of a semantics–pragmatics interface can be spelled out within Relevance Theory: words and sentences have a linguistic conceptual meaning that is only abstract or schematic (cf. sentences in an Austinian sense). The linguistic meaning gets enriched by words and sentences in use (cf. Austin's statements). Different theories tell a different story about the semantic enrichment of utterances (cf. the Recanati issue discussed above). Situation theorists exploit features of utterance and topic situation (see for example the work on unarticulated constituents initiated by Perry (1986), relevance theorists draw on relevance-driven pragmatic inferences

¹⁴ The most urgent follow-up problem concerning pointing and reference is how it is related to anaphora in discourse or dialog. This question is investigated in Poesio and Rieser (2009) using current dialog theory.

that are part of our theory of mind). Both approaches have in common that descriptive sentences *simpliciter* (linguistic meaning) cannot be said to be true or false. It is utterances of sentences (meaning in use, statements) that give rise to truth-bearing propositions; a sentence used in a certain utterance situation gets connected to the part of the world it is about – its “topic situation”. With regard to this connection the statement can be judged to correctly describe the topic situation or fail to do so (Austin, 1950; Barwise, 1989; Carston, 1998).¹⁵

Inherent in connecting an utterance α to the situation s it is about is to determine the referents (i.e. objects in s) of the nominators that are part of α . In its original formulation (Barwise and Perry, 1983), what linguistic nominal expressions refer to is fixed as a contextual parameter of a discourse situation d known as *speaker connections* c .¹⁶ c is a partial function from discourse situations and utterances onto objects available to the speaker.¹⁷ According to this line of thinking, reference and linguistic meaning might fall apart: a speaker connection can override the conventional meaning of a nominator (cf. Roberts, 1993:20). A similar view is maintained in RT (Powell, 2001b): a definite description “the F ” linguistically constrains (rather than determines) its interpretation to an individual concept of F in a salient context (Carston and Powell, 2006; Powell, 2001a). However, what is the salient context that gets exploited in the interpretation of nominators? Neither situation semantics nor RT gives a precise answer to this question. However, at this point one could plug-in Roberts (1993) account on “How Reference Works”. Roberts fuses philosophical reasoning with socio-psychological schemes and devises a perceptual model for referential expressions, the *figure-ground model*. The figure-ground model has three steps (Roberts, 1993:18–19).¹⁸

The very use of an indexical in relation to the context of the discourse determines the ground (e.g., the physical surroundings) which contains the referent. Gestures or actions which accompany the use of the indexical serve to narrow down the physical surroundings to a subsection which contains the referent. Descriptive content associated with the indexical term functions as a figure which makes the referent stand out in virtue of a contrast to the demonstrated segment of the physical surrounding.

The outcome of our study can be reconciled with the figure-ground model in a quite natural way, the delineating of the ground by means of a pointing cone can even be visualized. The pair consisting of a referring expression and a grounding action is construed as a meaning process: The pointing gesture highlights a certain subdomain, the referring expression picks out an object from this region. Neither the gesture nor the referring expression in isolation is as a rule capable of singling out a referent.

The remarkable feature of the role pointing is assigned to in Roberts’ framework is that it gets detached from reference. It is rather seen to be a “domain-fixer”, that is a semiotic means to delimit or direct attention to a salient context or situation or ground which in turn gives rise to the domain out of which the accompanying verbal description picks a referent. Thus, pointings are not likened to names, singular terms or definite descriptions, i.e., something that *eo ipso* refers (cf. the (Sem) hypotheses we started with), but to local adverbs that situate things or events. The change of focus is from “pragmatic reference” to the “location of objects” (cf. Kranstedt et al., 2007).

Despite the affinity of the conclusions we have drawn on empirical grounds to non-standard semantic–pragmatic theories, what is lacking in these paradigms, however, is the elaboration of the contribution of pointing to reference. The model developed in this article can thus serve as a foundation to the specifications made in situationist or relevance frameworks. In particular, it can spell out the “grounding” mechanism stipulated in the work of Roberts.

8. Conclusion

Using several philosophers’ (Wittgenstein’s, Kaplan’s, Davidson’s) and a great number of other cognitive scientists’ ideas about demonstrations accompanying verbal expressions, we set up hypotheses about pointing ranging from pure semantics to pragmatics. Pointing experiments based on body tracking, sophisticated measuring techniques and computer simulations involving a pointer’s distance to the object pointed at and his/her perspective to it have shown that a straight-forwardly defined pointing ray based notion of demonstration is easily falsified, especially for distal domains. A more intricate conceptualization of demonstration involving a pointing cone does not yield a unique demonstratum either but a set of objects covered by the base of the cone. The way out is to apply a heuristics which, however, uses the pointing cone’s main axis to predict the object pointed at.

¹⁵ The structure of a so-called Austinian proposition is ($s = \sigma$), where s is the situation a statement σ is about and s supports ($=$) σ .

¹⁶ Devlin (1991:227 *et seq.*), who provides the most thoroughly spelled-out picture in this regard, distinguishes between three types of speaker connections: “direct” (e.g., perceptual) connections, resource connections, and situation connections.

¹⁷ More formally, let OBJ_k denote the set of objects available to a speaker k , D the type of discourse situation – i.e. the set of all situations that satisfy the conditions that there is a speaker, a hearer and something spoken –, and α a nominator. Then $c : D \times \alpha \mapsto OBJ_k$.

¹⁸ Note that Roberts excludes proper names from his analysis.

As one major result of our empirical findings, we get that direct reference theories of demonstration are falsified by rigid data as are semantic or pragmatic theories using a naïve demonstration notion in the satisfaction condition for sentences or utterances. This certainly casts doubt on philosophers' theorizing concerning reference resting on the bias of easily individuated single objects.

A second major result pertains to the role of pragmatics in complex theories of grammar and theories of language. There, one leading text-book idea resting mainly on compositionality assumptions is to translate from a context-free syntax into a semantic representation interpreted in suitable models yielding the information of what is said and to add a pragmatic Grice-style component explaining things still open. If, however, already simple acts of demonstrative reference demand that we use heuristic inference, this picture can no longer be upheld. Instead, we arrive at a proposal using a pre-semantic pragmatics providing the demonstratum which can then enter into semantic composition proper. See also in this context the discussion of "reference by inference" throughout the paper. Incidentally, our findings also substantiate a proposal of Recanati's (see Recanati, 2004a) concerning the set-up of the semantics–pragmatics interface (see the highlighted passage on page 39 and our Fig. 16). Finally, we show that research in non-Tarskian, "Austinian" paradigms such as the language-as-situated activity program (a version of Situation Semantics), Relevance Theory, and the explanatory reference model of Roberts (1993) point towards a similar perspective and essentially support our conclusions. However, we need not go into alternative paradigms to demonstrate the ubiquitous role of pragmatics in comprehensive theories of language use.

Acknowledgments

Work on this paper has been supported by the German Research Foundation, in the CRC 360, *Situated Artificial Communicators*, and in the CRC 673, *Alignment in Communication*, both at Bielefeld University. The authors also want to thank the anonymous reviewers of the *Journal of Pragmatics* for their valuable comments. They helped to make the argument more stringent and to present the results in a more accessible way.

References

- Anscombe, Gertrude Elizabeth Margaret, von Wright, Georg Henrik (Eds.), 1969. *Ludwig Wittgenstein: Notebooks 1914–1916. Appendix I, Notes on Logic 1914–1916*. Basil Blackwell, Oxford.
- Austin, John L., 1950. *Truth*. In: *Proceedings of the Aristotelian Society. Supplementary*, vol. xxiv. Reprinted in John L. Austin, 1970. *Philosophical Papers, 2nd ed.*, Clarendon Press, Oxford, pp. 111–128.
- Bangerter, Adrian, 2004. Using pointing and describing to achieve joint focus of attention in dialogue. *Psychol. Sci.* 15 (6), 415–419.
- Bangerter, Adrian, Oppenheimer, Daniel M., 2006. Accuracy in detecting referents of pointing gestures unaccompanied by language. *Gesture* 6 (1), 85–102.
- Barwise, Jon, 1989. *The Situation in Logic*. CSLI Publications, Stanford, CA.
- Barwise, Jon, Perry, John, 1983. *Situations and Attitudes*. The David Hume Series of Philosophy and Cognitive Science Reissues. CSLI Publications, Stanford.
- Braun, David, 1994. Structured characters and complex demonstratives. *Philos. Stud.* 74 (2), 193–221.
- Braun, David, 1996. Demonstratives and their linguistic meanings. *Noûs* 30 (2), 145–173.
- Butterworth, George, 2003. Pointing is the royal road to language for babies. In: Kita, Sotaro (Ed.), *Pointing: Where Language, Culture, and Cognition Meet*. Lawrence Erlbaum Associates, Inc., Mahwah, NJ. pp. 9–33. (Chapter 2).
- Butterworth, George, Itakura, Shoji, 2000. How the eyes, head and hand serve definite reference. *Br. J. Dev. Psychol.* 18, 25–50.
- Carletta, Jean, 1996. Assessing agreement on classification tasks: the kappa statistic. *Comput. Linguist.* 2 (22), 249–254.
- Carston, Robyn, 1998. The semantics/pragmatics distinction: a view from relevance theory. *UCL Work. Pap. Linguist.* 10, 53–80.
- Carston, Robyn, Powell, George, 2006. Relevance theory – new directions and developments. In: Lepore, Ernest, Smith, Barry C. (Eds.), *The Oxford Handbook of Philosophy of Language*. Oxford University Press, Oxford, pp. 341–360.
- Chierchia, Gennaro, McConnell-Ginet, Sally, 2000. *Meaning and Grammar – An Introduction to Semantics*, 2nd ed. MIT Press, Cambridge.
- Clark, Herbert H., 1996. *Using Language*. Cambridge University Press, Cambridge.
- Clark, Herbert H., 2003. Pointing and placing. In: Kita, Sotaro (Ed.), *Pointing: Where Language, Culture, and Cognition Meet*. Lawrence Erlbaum Associates, Inc., Mahwah, NJ. pp. 243–269. (Chapter 10).
- Clark, Herbert H., Bangerter, Adrian, 2004. Changing ideas about reference. In: Noveck, Ira A., Sperber, Dan (Eds.), *Experimental Pragmatics, Palgrave Studies in Pragmatics, Language and Cognition*. Palgrave Macmillan, Basingstoke/New York. pp. 25–49. (Chapter 2).
- Cohen, Jacob, 1960. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* 20, 37–46.
- Cooperrider, Kensy, Núñez, Rafael, 2012. Nose-pointing: notes on a facial gesture of papua new guinea. *Gesture* 12 (2), 103–129. <http://dx.doi.org/10.1075/gest.12.2.01coo>
- Davidson, Donald, 1967. Truth and meaning. *Synthese* 17, 304–323.
- de Ruiter, Jan Peter, 2000. The production of gesture and speech. In: McNeill, David (Ed.), *Language and Gesture*. Cambridge University Press, Cambridge, UK. pp. 284–311. (Chapter 14).
- Devlin, Keith, 1991. *Logic and Information*. Cambridge University Press, Cambridge.
- Enfield, Nick J., 2001. Lip-pointing: a discussion of form and function with reference to data from laos. *Gesture* 1 (2), 185–212.

- Feinstein, Alvan R., Cicchetti, Domenic V., 1990. High agreement but low kappa: I. The problem of two paradoxes. *J. Clin. Epidemiol.* 43 (6), 543–549.
- Geurts, Bart, 1997. Good news about the description theory of names. *J. Semant.* 14 (4), 319–348. <http://dx.doi.org/10.1093/jos/14.4.319>
- Grice, Herbert Paul, 1975. Logic and conversation. In: Cole, Peter, Morgan, Jerry L. (Eds.), *Syntax and Semantics, Volume 3, Speech Acts*. Academic Press, New York, pp. 41–58.
- Gwet, Kilem, 2001. *Handbook of Inter-Rater Reliability*. STAXIS Publishing Company, Gaithersburg, MD.
- Harnish, Robert, 1991. Logical form and implicature. In: Steven, Davis (Ed.), *Pragmatics. A Reader*. Oxford University Press, Oxford. Reprinted from 1976, pp. 316–364. Reprinted from 1976.
- Haviland, John B., 2000. Pointing, gesture spaces, and mental maps. In: McNeill, David (Ed.), *Language and Gesture*. Cambridge University Press, Cambridge, MA, pp. 13–46.
- Kaplan, David, 1978a. On the logic of demonstratives. *J. Philos. Logic* 8, 81–98.
- Kaplan, David, 1978b. Dthat. In: Cole, Peter (Ed.), *Pragmatics, Volume 9 of Syntax and Semantics*. Academic Press, New York/San Francisco/London, pp. 221–243.
- Kaplan, David, 1989a. Demonstratives. In: Almog, Joseph, Perry, John, Wettstein, Howard (Eds.), *Themes from Kaplan*. Oxford University Press, New York and Oxford, pp. 481–563.
- Kaplan, David, 1989b. Afterthoughts. In: Almog, Joseph, Perry, John, Wettstein, Howard (Eds.), *Themes from Kaplan*. Oxford University Press, New York/Oxford, pp. 565–614.
- Kendon, Adam, 1980. Gesticulation and speech: two aspects of the process of utterance. In: Key, Mary Ritchie (Ed.), *The Relationship of Verbal and Nonverbal Communication, Volume 25 of Contributions to the Sociology of Language*. Mouton, The Hague, pp. 207–227.
- Kendon, Adam, 1997. *Gesture*. *Annu. Rev. Anthropol.* 26, 109–128.
- Kendon, Adam, 2000. Language and gesture: unity or duality? In: McNeill, David (Ed.), *Language and Gesture*. Cambridge University Press, Cambridge, UK, pp. 47–63. (Chapter 2).
- Kendon, Adam, 2004. *Gesture: Visible Action as Utterance*. Cambridge University Press, Cambridge, MA.
- Kendon, Adam, Versante, Laura, 2003. Pointing by hand in “neapolitan”. In: Kita, Sotaro (Ed.), *Pointing: Where Language, Culture, and Cognition Meet*. Lawrence Erlbaum Associates, Inc., Mahwah, NJ, pp. 109–137. (Chapter 6).
- Kita, Sotaro (Ed.), 2003. *Pointing: Where Language, Culture, and Cognition Meet*. Lawrence Erlbaum Associates, Inc., Mahwah, NJ.
- Kjærsgaard-Andersen, Per, Christensen, Finn, Schmidt, Steen Asmus, Pedersen, Niels Wisbech, Bent, Jørgensen, 1988. A new method of estimation of interobserver variance and its application to the radiological assessment of osteoarthritis in hip joints. *Stat. Med.* 7 (6), 639–647. <http://dx.doi.org/10.1002/sim.4780070603>
- Kranstedt, Alfred, Lücking, Andy, Pfeiffer, Thies, Rieser, Hannes, Staudacher, Marc, 2006. Measuring and reconstructing pointing in visual contexts. In: Schlangen, David, Fernández, Raquel (Eds.), *Proceedings of the 10th Workshop on the Semantics and Pragmatics of Dialogue, Brandial’06*. Universitätsverlag Potsdam, Potsdam, pp. 82–89.
- Kranstedt, Alfred, Lücking, Andy, Pfeiffer, Thies, Rieser, Hannes, Wachsmuth, Ipke, 2006b. Deixis: how to determine demonstrated objects using a pointing cone. In: Gibet, Sylvie, Courty, Nicolas, Kamp, Jean-Francois (Eds.), *Gesture in Human-Computer Interaction and Simulation*. Springer, Berlin, pp. 300–311.
- Kranstedt, Alfred, Lücking, Andy, Pfeiffer, Thies, Rieser, Hannes, Wachsmuth, Ipke, 2006c. Deictic object reference in task-oriented dialogue. In: Rickheit, Gert, Wachsmuth, Ipke (Eds.), *Situated Communication*. Mouton de Gruyter, Berlin, pp. 155–207.
- Kranstedt, Alfred, Lücking, Andy, Pfeiffer, Thies, Rieser, Hannes, Staudacher, Marc, 2007. Locating objects by pointing. In: *3rd International Conference of the International Society for Gesture Studies*, Evanston, IL, USA.
- Krauss, Robert M., Chen, Yihsiu, Gottesmann, Rebecca F., 2000. Lexical gestures and lexical access: a process model. In: McNeill, David (Ed.), *Language and Gesture*. Cambridge University Press, Cambridge, UK, pp. 261–283. (Chapter 13).
- Kripke, Saul A., 1980. *Naming and Necessity*. Blackwell, Oxford.
- Kühnlein, Peter, Stegmann, Jens, 2003. Empirical issues in deictic gestures: Referring to objects in simple identification tasks. Technical Report 2003/03, SFB 360. Universität Bielefeld.
- Lepore, Ernest, Ludwig, Kirk, 2007. *Donald Davidson’s Truth-Theoretic Semantics*. Oxford University Press, Oxford, NY.
- Levelt, Willem J.M., 1989. *Speaking: From Intention to Articulation*. MIT Press, Cambridge, MA.
- Levinson, Stephen C., 1983. *Pragmatics*. Cambridge University Press, Cambridge.
- Levinson, Stephen C., 2008. Deixis. In: *The Handbook of Pragmatics*. Blackwell, pp. 97–121. (Chapter 5).
- Lücking, Andy, Stegmann, Jens, 2005. Assessing reliability on annotations (2): statistical results for the DEIKON scheme. Technical Report 3, SFB 360. Universität Bielefeld.
- Lücking, Andy, Rieser, Hannes, Stegmann, Jens, 2004. Statistical support for the study of structures in multi-modal dialogue: inter-rater agreement and synchronization. In: *Proceedings of the Eighth Workshop on the Semantics and Pragmatics of Dialogue, Catalog’04*, Barcelona, pp. 56–63.
- Lyons, John, 1977. *Semantics, vol. 2*. Cambridge University Press, Cambridge, UK.
- McGinn, Colin, 1981. The mechanism of reference. *Synthese* 49 (2), 157–186.
- McNeill, David, 1992. *Hand and Mind – What Gestures Reveal about Thought*. Chicago University Press, Chicago.
- McNeill, David (Ed.), 2000. *Language and Gesture*. Cambridge University Press, Cambridge, UK.
- Pechmann, Thomas, Deutsch, Werner, 1982. The development of verbal and nonverbal devices for reference. *J. Exp. Child Psychol.* 34 (2), 330–341. [http://dx.doi.org/10.1016/0022-0965\(82\)90050-9](http://dx.doi.org/10.1016/0022-0965(82)90050-9)
- Peirce, Charles Sanders, 1867. On a new list of categories. In: *Proceedings of the American Academy of Arts and Sciences Series, vol. 7*. pp. 287–298.
- Peirce, Charles Sanders, 1965. *Collected Papers of Charles Sanders Peirce*. Repr. from 1932.
- Perry, John, 1986. Thought without representation. In: *Supplementary Proceedings of the Aristotelian Society, vol. 60*. pp. 263–283.
- Perry, John, 2009. Directing intentions. In: Almog, Joseph, Leonard, Paolo (Eds.), *The Philosophy of David Kaplan*. Oxford University Press, Oxford, NY, pp. 187–207. (Chapter 11).

- Pfeiffer, Thies, 2011. *Understanding Multimodal Deixis with Gaze and Gesture in Conversational Interfaces* Shaker Verlag, Bielefeld University, Aachen (Ph.D. thesis).
- Pfeiffer, Thies, 2012. Interaction between speech and gesture: strategies for pointing to distant objects. In: Eftimiou, Eleni, Kouroupetroglou, Georgios, Fotinea, Stavroula-Evita (Eds.), *Gestures and Sign Language in Human–Computer Interaction and Embodied Communication*, 9th International Gesture Workshop, GW 2011, number 7206 in LNAI, May 2011, Athens, Greece. Springer-Verlag GmbH, pp. 238–249. <http://dx.doi.org/10.1007/978-3-642-34182-3>
- Pfeiffer, Thies, Kranstedt, Alfred, Lücking, Andy, 2006. *Sprach-Gestik Experimente mit IADE, dem Interactive Augmented Data Explorer*. In: *Dritter Workshop Virtuelle und Erweiterte Realität der GI-Fachgruppe VR/AR, Koblenz*, pp. 61–72.
- Poesio, Massimo, Rieser, Hannes, 2009. *Anaphora and direct reference: empirical evidence from pointing*. In: *Proceeding of the 13th Workshop on the Semantics and Pragmatics of Dialogue, DiaHolmia*, pp. 35–42.
- Powell, George, 2001a. *Complex demonstratives*. *UCL Work. Pap. Linguist.* 13, 43–71.
- Powell, George, 2001b. *The referential-attributive distinction – a cognitive account*. *Pragmat. Cogn.* 9 (1), 69–98.
- Quine, Willard Van Orman, 1960. *Word and Object*. MIT Press, Cambridge, MA.
- Recanati, François, 2004a. *Literal Meaning*. Cambridge University Press, Cambridge, MA.
- Recanati, François, 2004b. *Pragmatics and semantics*. In: Horn, Laurence R., Ward, Gregory (Eds.), *The Handbook of Pragmatics*. Blackwell, Malden, MA. pp. 442–463. (Chapter 20).
- Roberts, Lawrence D., 1993. *How Reference Works: Explanatory Models for Indexicals, Descriptions, and Opacity*. State University of New York Press, Albany.
- Russell, Bertrand, 1905. *On denoting*. *Mind* 14 (56), 479–493.
- Sperber, Dan, Wilson, Deirdre, 1986. *Relevance: Communication and Cognition*. Blackwell, Oxford.
- Stegmann, Jens, Lücking, Andy, 2005. *Assessing reliability on annotations (1): Theoretical considerations*. Technical Report 2, SFB 360. Universität Bielefeld.
- Tarski, Alfred, 1935. *Der Wahrheitsbegriff in den formalisierten Sprachen*. *Stud. Philos.* 1, 261–405.
- Tarski, Alfred, 1944. *The semantic conception of truth and the foundations of semantics*. *Philos. Phenomenol. Res.* 4, 341–375.
- van der Sluis, Ielka, Kraemer, Emiel, 2007. *Generating multimodal references*. *Discourse Process.* 44 (3), 145–174. <http://dx.doi.org/10.1080/01638530701600755>, Special Issue on "Dialogue Modelling: Computational and Empirical Approaches".
- Vanderveken, Daniel, 1990. *On the unification of speech act theory and formal semantics*. In: Cohen, Philip R., Morgan, Jerry L., Pollack, Martha E. (Eds.), *Intentions in Communication*. MIT Press, Cambridge, MA. pp. 195–220. (Chapter 11).
- Zeevat, Henk, 1999. *Demonstratives in discourse*. *J. Semant.* 16 (4), 279–313. <http://dx.doi.org/10.1093/jos/16.4.279>