Center for
Mathematical Economics
Working Papers

519

July 2014

Categorization Based Belief Formations

Jörg Bleile

Center for Mathematical Economics (IMW)
Bielefeld University
Universitätsstraße 25
D-33615 Bielefeld · Germany

e-mail: imw@uni-bielefeld.de
http://www.imw.uni-bielefeld.de/wp/
ISSN: 0931-6558

Categorization based Belief formations

Jörg Bleile*

July 21, 2014

Abstract

An agent needs to determine a belief over potential outcomes for a new problem based on past observations gathered in her database (memory). There is a rich literature in cognitive science showing that human minds process and order information in categories, rather than piece by piece. We assume that agents are naturally equipped (by evolution) with a efficient heuristic intuition how to categorize. Depending on how available categorized information is activated and processed, we axiomatize two different versions of belief formation relying on categorizations. In one approach an agent relies only on the estimates induced by the single pieces of information contained in so called target categories that are activated by the problem for which a belief is asked for. Another approach forms a prototype based belief by averaging over all category-based estimates (so called prototypical estimates) corresponding to each category in the database. In both belief formations the involved estimates are weighted according to their similarity or relevance to the new problem. We impose normatively desirable and natural properties on the categorization of databases. On the stage of belief formation our axioms specify the relationship between different categorized databases and their corresponding induced (category or prototype based) beliefs. The axiomatization of a belief formation in Billot et al. (Econometrica, 2005) is covered for the situation of a (trivial) categorization of a database that consists only of singleton categories and agents basically do not process information categorical.

Keywords: Belief formation, prior, case-based reasoning, similarity, categorization, prototype.

JEL Classification: D81, D83

*Center for Mathematical Economics (IMW) and Bielefeld Graduate School of Economics and Management (BiGSEM), Bielefeld University, jbleile@uni-bielefeld.de.
This research was carried out within the International Research Training Group "Economic Behavior and Interaction Models" (EBIM) financed by the German Research Foundation (DFG) under contract GRK 1134/2. Financial support by the IMW is gratefully acknowledged.

1 Introduction

Often agents need to evaluate and judge the likelihood of future uncertain events. On which basis can individuals derive and assign likelihoods and form probabilistic beliefs over random incidents?

Traditionally, economic theory models uncertainties in a state space representation a la Savage (1954) and Bayes and derive a subjective prior based on observable actions of the agent. However, this procedure implicitly assumes that agents already know or are endowed with a subjective prior belief, which they express through their observable actions. In this way, the Savage and Bayesian approach does not advice agents how to find or form a prior explicitly. Basically, the belief is purely subjective and offers no mechanism to incorporate information directly into a belief formation. Consequently, their normatively appealing and convincing approach to endogenously derive a belief is not feasible in situation in which an agent might not be able to condense her insufficient or too complex information into a consistent state space.

We consider an axiomatization of belief formation that allows and requires to take directly into account the available information (gathered in form of a list or database of past observations or cases). The influence of data and experience on the formation of a probabilistic belief was examined initially by the axiomatization of Billot et al. (2005) (BGSS from now on). The axiomatizations of BGSS and related ones of Eichberger and Guerdjikova (2010) (EG) (for ambiguous multiprior beliefs) and Bleile (2014a) (precision dependent cautious beliefs)[1] yield that a belief induced by a database is a similarity weighted average of the estimations induced by all observed cases in the database. Thereby similarity weights capture different degrees of relevance of the potentially very heterogenous information.

A common shortcoming of these approaches to belief formation is that an agent processes each distinct single piece of information separately and forms its induced estimate. Interpreting a database as memory an agent is assumed to store (memorize) all single pieces of information and needs to retrieve any single piece of information from her memory.

However, numerous studies in (social) psychology and cognitive science show that humans do not store and treat single pieces of information in such a one by one procedure, but classify information in different categories. The prominent social psychologist Allport (1954) memorably noted "the human mind must think with the aid of categories. We cannot possibly avoid this process. Orderly living depends upon it ". There is a wealth of research demonstrating that humans' cognition processes information by employing categorical thinking, reasoning and stereotyping.[2] In particular, one can interpret categorization as model of similarity-based reasoning (Tversky (1977), Gilboa and Schmeidler (1995,2001)) in which information needs not to be understood in its particularity, but as member of a larger classified category that allows to generalize properties from cate-

[1]Another axiomatization that does not take into account all potentially available information in this vein is Bleile (2014b). It deals with a two stage belief formation that consists of a initial filtering process that "screens and selects" the information that finally flows into the belief formation process.

[2]The psychology literature on categorization is vast, e.g. see Rosch and Lloyd (1976), Murphy and Medin (1985), Goldstone (1994), Rips (1989), Smith et al. (1998), Medin and Aguilar (1999), Murphy (2002). Real life examples discuss that consumers categorize products (Smith 1965), investors engage in "style investing, (Sharpe (1992), Bernstein (1995)), rating agencies categorize firms wrt. default risk (Coval et al. 2009), etc.

gory members to new members through analogies and similarities. This makes categorical thinking especially helpful for predictions (Osherson et al. (1990), Anderson (1991)).

In order to capture the impact of categorical thinking and reasoning in agent's belief, we modify and extend the mentioned axiomatic approaches (in particular BGSS) by adding a categorization procedure that affects the processing, storing, retrieving and employing of potentially available information.

In complex and poorly understood environments, categorizations emerge naturally to simplify actions by gathering many distinct experiences together and ignoring the details of each single piece of information. Limited learning and memorizing opportunities drive agents into relying on abstractions and (categorical) summarizations rather than on single past cases. Processing and storing of all past cases in full detail bears costs in storing and retrieving the information, since the finer information is stored the more effort is required to activate it. The classification of information in different categories offers a less demanding way of storing and retrieving information, since only the assignment to suitable categories and their characteristics needs to be memorized. In particular, the literature on "optimal" categorization focuses on the issue how fine or coarse categories ought to be formed in order to process information in a way to gain a maximum amount of information with the least cognitive effort. In particular, it should be more efficient than some other form of case-based reasoning, as for instance kernel-based estimation.

Another important function of categorical reasoning concerns its role for facilitating and improving inductive inference and prediction. The underlying idea is that an assignment to categories does allow an agent not only to use the information contained in the current problem, but exploit as well the additional information provided by the categories to which this problem belongs (or which it activates). Of course, this is only helpful if the previous experiences contained in the specific categories provide some information for the actual problem such that the agent can infer or generalize some information and properties from past observations in the categories. In this spirit, categorization is also closely related to the ideas of reasoning by similarities or analogies. From this perspective, a categorization of information enables and implicitly provides an agent with additional (more detailed) information than mentioned in the initial description of the problem. Ideally categories are formed like sufficient statistics for its assigned members and thus would make prediction particularly simple and reliable.[3]

In this paper, we are not concerned with the formation of categories, but assume that a set of (optimal) categories is already naturally or evolutionary determined.[4] In particular, we are solely interested in an axiomatic description on how categorized information is incorporated into a belief formation by agents.

The categorization literature identified several procedures in using categorical thinking for belief formation. The approaches differ in the way how many categories are taken into account. Either all categories are considered or only some specific target category(ies) are taken into account. Another difference concerns (the still an ongoing discussion about)

[3]Peski's (2011) categorization model can be interpreted as such an optimal statistical procedure.

[4]Traditionally, categories are formed based on (attribute-wise, overall, functional or casual) similarity considerations. Roughly speaking, in general categories are often formed as to maximize the similarity of objects within a category and the dissimilarity of objects from different categories. However, there is an ongoing discussion and debated whether categorization presupposes a notion of similarity or not (see Goldstone (1994) and Gärdenfors (2000), Pothos (2005)). For instance some literature argue that categorization is theory or rule-based (according to various criteria)

the issue how categories are represented themselves. Either categories are represented by an aggregated summarizing representative, that captures the essence or central tendency of the category - a so called prototype - or all members of the category are used for its representation.[5]

There is experimental evidence in psychology that individuals tend to rely on (a single) most likely target category(ies), whereas the other categories (and their content) are immaterial for the belief formation (e.g. Murphy and Ross (1994), Krueger and Clement (1994), Malt et al. (1995)). When faced with a new problem, an agent's mind activates automatically some already generated category(ies) that are best fitting according to some metric for the current problem.[6] Depending on how an agent treats categories she will form her belief either based on all single pieces of information contained in the target category(ies) or use the estimates induced by a prototypical representative associated with the (target) category(ies). Our first axiomatization of a category based belief formation will adopt this approach based on activated target categories, which simplifies (cognitively) the belief formation, since an agent only needs to process the information that is directly evoked for the current problem.

The second stream of literature -which is covered in our second axiomatization of a prototype based belief- is based on the prototype of all categories in a database. Such a prototype based belief adopts the approach taken in Anderson (1991) and models the situation, in which an agent might not be able to figure out best fitting target category(ies). The simplifying power of categorization in this approach results from taking into account all (categorized) information through prototypical summaries and not by memorizing and retrieving all single pieces of information. This reduces the cognitive and memorization effort. It averages the particular prototypical estimates induced by the categories.

Our axiomatizations of both kinds of belief formation -category and prototype based- are based on modified and extended versions of the axioms of BGSS and partly Bleile (2014b). Categorization based belief formations can be seen as two stage procedures. First, agents are endowed with a natural categorization structure on the in principle available information (e.g. like through a natural or evolutionary developed optimal heuristic algorithm). On the basis of this categorization structure and the current problem, the available information (provided by the database/memory of past experiences) will be categorized. We will assume some reasonable, natural and well known -but rather weak- properties on this induced categorization of databases. The (structural) properties on the (induced) categorization of databases differ for our two versions of categorization based belief formation. However, a common feature concerns our main requirement that an agent will not categorize any database, but that databases must be sufficiently complex or diverse to initiate a categorization process. We assume that a database need to contain a minimum amount of distinct cases such that an agent really (wants to) thinks in categories.[7] However, this is a quite natural requirement, since one reason for categorizing information is to overcome limitations in processing cognitively challenging information or environments.

[5] There are also approaches in between (Vanpaemel and Storms (2008)), but we stick to the extreme cases.

[6] For example by comparing the actual problem to the prototypical problem of different categories until a closest match is found. The automaticity in categorical thinking is discussed e.g. in Allport (1954), Bargh (1994, 1997, 1999)

[7] Complexity is certainly related to the number of options to be considered, but also few options characterized by difficult interwoven features might be challenging to evaluate.

Another common property says that the order of cases in a database is immaterial for the categorization, i.e. a categorization procedure depends only on content and not on the sequences of pieces of information.
For a prototype based belief we require that a minimum number of distinct cases a "real" category ought to contain. In particular, we require for a prototype based belief formation that a database is categorized (in some accordance with the natural categorization) in such a way that at most one singleton category exits and all other non empty categories consist of at least two members. Of course a degenerate categorization in which each category contains only one member is meaningless for our purpose (and is covered directly by BGSS).

The second stage of a belief formation based on categorized information deals with the behavioral impacts on the actual belief level. As in BGSS, we require that the belief is independent of the order of the (categorized) information and that some form of concatenation Axiom holds. Explicitly, a belief induced by the combination of two databases should be a weighted average of the beliefs induced by the two databases separately. However to keep the normatively appealing spirit of the axiom, we need to ensure that the categorized information of the combination of two databases coincides with the combination of the two separately categorized underlying databases.[8]

The particular properties of the categorization procedures and the axioms on the belief formation level guarantee that the beliefs based on categorized information can be represented only based on information in the target categories evoked by the new problem (i.e. as the category based belief) or based only on all prototypes of the categorized information (i.e. as the prototype based belief).

As already mentioned there exist research discussing predictions based on categorical thinking, but none of them is of an axiomatic nature. However, we are solely interested in the behavioral foundation of a belief formation based on categorized information. Our approach is closest to BGSS and Bleile (2014b) with regard to axiomatizations of belief formations. Concerning the axiomatized procedures, the most relevant works are Anderson (1991) for the prototype based belief and Murphy and Ross (1994) for the category based belief. The existing literature deals either with applying prediction procedures based on categorized information and discussing its consequences for specific situation (e.g. Mullainathan (2002)) or is concerned with how and why "optimal " formations of categories emerge (Fryer and Jackson (2008), Mohlin (2014), Peski (2011)). The mentioned papers all employ a belief formation relying on prototypes (of the main target categories or the category the current problem belongs to). However they all differ with regard to their notion of an "optimal" categorization. In Fryer and Jackson (2008) the optimal categorization minimizes the sum (across categories) of within category variations between objects that have already been encountered (for an exogenously fixed number of categories). Mohlin (2014) aims to find the optimal number of categories in order to minimize the prediction error, which amounts to tackling the tradeoff between small and fine or more coarse but larger categories to avoid overfitting problems (which is also the principle rational of Peski (2011) and Al Najjar and Pai (2014) for decision making). Peski (2011) shows when cat-

[8]For the axiomatization of a prototype based belief we deal with this issue in a similar spirit as in Bleile (2014b) by introducing another simultaneously available (super)-database, which serves as the common reference which "dictates" the categorization in a consistent manner for any sub-database.

egorical learning is optimal for prediction (in the sense of an (asymptotic) statistical tool equivalent to learning by Bayesian updating). He compares a categorization algorithm with Bayesian updating. The underlying assumption is that the environment is symmetric meaning that the Bayesian prior is symmetric. The categorization algorithm is such that categories are formed in order to minimize the inner entropy of the categories, i.e. to maximize the informational content in the categories. His categorization procedure combines deductive reasoning (i.e. learn to form categories) and applying the deduced categorization inductively to belief formation. This is structural impossible in our axiomatization.

The following section will introduce the database related framework. Section 3 illustrates by means of an example the two beliefs based on categorizations. In section 4 our natural categorization structure is discussed. Section 5 and 6 cover the categorization of databases, the axioms on the belief level and the resulting belief formations for both categorization based belief formations separately. We conclude in the last section. All proofs can be found in the last section.

2 The model

In this section, we introduce the case-based information framework and the basic building blocks of our belief formation based on filtered information. Further, we introduce some definitions and notations necessary for our approach.

2.1 Database framework

A basic case $c = (x, r)$ consists of a description of the environment or problem $x \in X$ and an outcome $r \in R$, where $X = X^1 \times X^2 \times .... \times X^N$ is a finite set of all characteristics of the environment, in which X^j denotes the set of possible values features j can take. R denotes a finite set of potential outcomes, $R = \{r^1, ..., r^n\}$

The **ordered** set $C \subseteq X \times R$ consists of all $m \geq 3$ basic cases, i.e. $C = \{c^1, ..., c^m\}$.

A database D is a sequence or list of basic cases $c \in C$. The set of databases D consisting of L cases, i.e. $D = (c_1, ..., c_L)$ where $c_i \in C$ for all $i \leq L$, is denoted by C^L and the set of all databases by $C^* = \cup_{L \geq 1} C^L$, including the empty database $\emptyset$. The description of databases as sequence of potentially identical cases allows multiple observation of an identical case to be taken into account and treated as an additional source of information. For a database $D \in C^*$, $f_D(c)$ denotes the relative frequency of case $c \in C$ in databases D. The concatenation of two databases $D = (c_1, c_2, ..., c_L) \in C^L$ and $E = (c'_1, c'_2, ..., c'_T) \in C^T$ (where $c_i, c'_j \in C$ for all $i \leq L, j \leq T$) is denoted by $D \circ E \in C^{L+T}$ and is defined by $D \circ E := (c_1, c_2, ..., c_L, c'_1, c'_2, ..., c'_T)$.

In the following we will abbreviate the concatenation or replication of L-times the identical databases D by D^L. Specifically, c^L represents a database consisting of L-times case c.

For any $D \in C^*$ the diversity of a database D is given by $div(D) := |\{D\}|$, where as usual $\{D\}$ denotes the set of different cases contained in database D. So $div(D)$ gives the number of different cases contained in database D.

We need to translate some relations from sets to the list framework.

Definition 2.1

(i) The $\in$-relation on databases is defined by $c \in D$ if $f_D(c) > 0$.

(ii) The $\subseteq$-relation on the set of databases C^* is defined by $D \subseteq E \Leftrightarrow f_D(c)|D| \leq f_E(c)|E|$ for all $c \in C$. We will call such databases to be nested.
(iii) The $\cap$-relation on databases is given by $D \cap E = ((c^{\min\{f_D(c)|D|, f_E(c)|E|\}})_{c \in C})$
(iv) Two databases D and E are disjoint if for all $c \in C$: $c \in D$ if and only if $c \notin E$.

The definitions are basically independent of the order of cases in the databases. Note however that the definition of $\cap$-relation in (iii) is very specific, since the order of C is transferred, i.e. by intersection a specific order (on C) is induced.[9]

2.2 Induced belief

For a finite set S, $\Delta(S)$ denotes the simplex of probability vectors over S and for $n \in \mathbb{N}$ Δ^n denotes the simplex over the set $\{1, 2, ..., n\}$.
As in BGSS, EG and Bleile (2014a) an agent forms a belief over outcomes $P(x, D) \in \Delta(R)$ in a certain problem characterized by $x \in X$ using her information captured in a database $D \in C^*$, i.e. $P : X \times C^* \to \Delta(R)$.

3 Motivating example

In order to illustrate the basic idea and plausibility of categorization based belief formation, we incorporate the two categorization procedures into the doctor example of BGSS.

A doctor needs to evaluate different outcomes of a treatment. She has some working experience or access to some medical database related to the treatment $D = (c_1, ..., c_l)$, where she recorded in a case $c_i = (x_i, r_i)$ the vector of characteristics of a patient i, $x_i \in X$, (e.g. age, gender, weight, blood count, specific illness) and the observable outcome of the treatment $r_i \in R$ (e.g. better, worse, adverse effects).
A new patient characterized by x enters her office and using a medical record D, the doctor wants to derive a belief $P_x(D) \in \Delta(R)$ over potential outcomes in R. She might apply an empirical frequency and use only a part D_x of the database D, which contains only cases $c = (x_c, r_c)$ of patients with *"identical"* characteristics $x_c = x$ compared to the current patient,

$$\text{"Frequentist":} \qquad P_x(D) = \frac{\sum_{c \in D_x} \delta_{r_c}}{|D_x|}$$

However, if the database contains not sufficiently many of these *"identical"* patients x, she might want to include also *"similar"* patients. She judges the degree of similarity between patients x and x' by $s(x, x') \in \mathbb{R}_+$. Further, she might induce from a case $c = (x_c, r_c)$ not only a point estimate δ_{r_c} on the realized outcome, but derives a more general estimate $P^c \in \Delta(R)$ on likelihoods of particular (related) outcomes and forms the belief as axiomatized in BGSS (2005) by

$$\text{"BGSS-belief":} \qquad P_x(D) = \frac{\sum_{c \in D} s(x, x_c) P^c}{\sum_{c \in D} s(x, x_c)}$$

[9] In contrast to intersections of sets, where orderings are immaterial, intersection of databases do require some assumption on resulting orderings. Alternatively, one also might keep the sequence of either database.

However, as discussed above, the cognitive science literature emphasized the role of categories in storing, retrieving and processing of information. Also the literature argued for a naturally (by evolution) given ability or heuristic feeling to categorize. Thus, assume the doctor is implicitly able to categorize the set C of all potentially possible patient-outcome pairs $(x, r) \in C$ into different categories $\tilde{C}_l \subset C$ for $l \leq L$, such that the set $\tilde{C} = \{\tilde{C}_1, .., \tilde{C}_L\}$ partitions the set of all possible cases $c = (x, r)$. For example category $\tilde{C}_1$ contains all male patients, with age below 60 years, any weight, good blood count and sore throats and category $\tilde{C}_l$ contains all male patients, overweight and heart problems and so forth. In general, categories might be exclusive or non-disjoint, for instance male patients might appear in different categories. Such an implicit preexisting natural categorization structure might be subconsciously rooted in the mind of a doctor and induces some consistent (embedded) categorization of a database D. Alternative, one might think of the preexisting natural categorization structure as the patient groups that the doctor are taught at medical school and the database D as the experience she made in working in a hospital. Receiving information about a new patient, she wants to predict the outcome based on (a natural or taught) categorization of her past experience. Depending on how the doctor uses her categorized patient database D, she might form a category or prototype based belief as follows.

Suppose the current patient x is male, older than 40 years, overweight, etc. and the doctor knows for each characteristic its value. The doctor partitioned her experience D with the treatment into categories consistent with the natural (preexisting) subconscious categorization. A specific patient x might trigger, activate or evoke automatically the category(ies) to which this patient belongs, is related to or matches best. For example this might be the category $\tilde{C}_1$ (as above) "synchronized"/intersected with the actually experienced database D, which we denote by $\tilde{C}(x, D) \subseteq D$. Thus, not only the identical patient profiles are recalled and considered but the entire x-evoked category in D that also may contain different, but somehow similar or related patients (according to the criteria for categorization, which were assumed to be optimal for predictive tasks by evolution). A doctor might form a **category based belief** based only on the members the category(ies) that patient x evokes, i.e. $\tilde{C}(x, D)$,

$$\text{"Category based belief":} \qquad P_x(D) = \frac{\sum_{c \in \tilde{C}(x,D)} s(x, x_c) P^c}{\sum_{c \in \tilde{C}(x,D)} s(x, x_c)},$$

where as above s measures the similarity between the current patient and the already treated patients.

Of course $\tilde{C}(x, D)$ might not consist only of a single category, but several categories that are activated by the patient x.

Alternatively, the doctor might have already categorized her experience D prior a new patient arrives, i.e. $\tilde{C}^D = \{\tilde{C}_1^D, .., \tilde{C}_L^D\}$ for $l \leq L$, such that $\bigcup_{l \leq L} \tilde{C}_l^D = D$ in a consistent manner with respect to the (preexisting) natural categorization structure $\tilde{C}$. Furthermore, she might have formed some prototypical estimates $P^{\tilde{C}_l^D} \in \Delta(R)$ for each of these categories. For a given database D and a new patient x, the doctor's **prototype based belief** in such a situation might be given by a weighted average of these prototypical estimates, where the weights are determined by the relevance of the particular category for the current patient, i.e. $\tilde{s}(x, \tilde{C}_l^D)$ for all $l \leq L$,[10]

$$\text{"Prototype based belief":} \qquad P_x(D) = \frac{\sum_{l \leq L} \tilde{s}(x, \tilde{C}_l^D) P^{\tilde{C}_l^D}}{\sum_{l \leq L} \tilde{s}(x, \tilde{C}_l^D}$$

4 Natural evolutionary (optimal) categorization structure

Definition 4.1

A (natural) categorization structure $\tilde{C} = \{\tilde{C}_1, .., \tilde{C}_L\}$ on the set of basic cases C partitions C into L different nonempty categories $\tilde{C}_l \subseteq C$ for $l \leq L \in \mathbb{N}$, i.e. $C = \bigcup_{l \leq L} \{\tilde{C}_l\}$.

The definition allows for non-disjoint categories, since it is quite naturally that a case could be classified into multiple categories if a categorization depends on more than one criterion. For instance, the categories of young and male patients are not necessarily disjoint, when gender and age are criteria. Moreover, hierarchical categories are not mutually exclusive, e.g. the category of young patients and the one of young male patients.

We will not care about the formation of categories and assume that agents are naturally endowed with an idea, how to construct categories. In particular, through evolutionary pressure nature equipped us with an heuristic algorithm or tool that allows to form categorizations that organize our experience in an almost optimal way for prediction and that tends to minimize prediction errors (and thus increase the likelihood to survive and stay fit. concerning fitness and survival). This is supported by the findings that young children appear to form, acquire and use categories from very early on (Gelman and Markman (1986), Smith (1989), Murphy (2002)), showing also that many categorizations are innate.[11] However, if the categorization is based on data, the developmental literature shows that especially in the beginning of learning (i.e. with small databases (i.e. for children)) the categorization is still flexible (Hayne (1996), Quinn and Eimas (1996)). This concern is immaterial for a natural categorization, since it is based already on all potential pieces of information. However, we will take care of it, when we consider categorization induced by database (in particular in the framework of a prototype based belief formation).

Another justification for assuming such a fixed preexisting categorization structure on the set of all potential cases is by interpreting it as a result of an already developed optimal categorization with regard to numbers and content of the categories. The literature varies in the way how they define an optimal categorization (as already discussed in the introduction, e.g. Fryer and Jackson (2008), Peski (2011), Mohlin (2014)).

A more direct reason, why we assume a preexisting natural categorization is based on the fact that we want to avoid the many difficult and interacting mechanisms involved in a categorization process[12] and want to focus solely on the belief formation issue.

Also in the literature it is not uncommon to assume a fixed preexisting categorization structure (Anderson (1991), Murphy and Ross (1994), Mullainathan (2002), Manzini and

[10] Alternatively, $\tilde{s}$ measures likelihood of patients x to be assigned or belonging to the particular category

[11] It is not observable that children rely on on purely empirical learning.

[12] e.g. initial encoding, abstraction of conceptual representation (if any), storage in memory of the abstraction and/or exemplars, retrieval of stored representations, decision process that produce categorization or typicality, see Murphy and Ross (1994)

Mariotti (2012), Al-Najjar and Pai (2014) or Mohlin's (2014) ex ante optimal categorization).

5 Axiomatization of category based belief formation

5.1 Specific properties of the categorization procedure

In this section we specify the list of categories a problem $x \in X$ evokes or activates and our concept of a x-evoked categorization of a database.

Problem evoked categorization of a database

The definition of a problem evoked categorization of information captures the intuition that a new problem activates specific most appropriate or relevant categories (that are already generated in the natural categorization structure). An agent takes into account these "target" categories for the current problem.

Definition 5.1
For all $x \in X$ a categorization structure $\tilde{C}$ induces a list $\tilde{C}_x$ of categories that problem x evokes. For all x there exist a $M_x \subseteq \{1, .., L\}$, such that $\tilde{C}_x := ((\tilde{C}_l)_{l \in M_x}) \in C^*$
We call $\tilde{C}_x$ the ***categories that are activated (evoked) by problem x or short x-activated categories****.*

There is substantial experimental evidence showing that when faced with an object, humans' brains automatically activate category(ies) that (according to some metric) appears to suit the current problem best (with regard to best fitting, most likely or analogous category(ies)).[13] Basically, a new problem does not trigger some most relevant single pieces of information, but the activation process is based on categorical thinking.
Our definition does not specify the exact procedure of activation. However, for our purpose to form a belief given a current problem , i.e. the characteristic part of case $x \in X$, it is most reasonable to think about a categorization of past observations according to their characteristics.[14] In this way, the characteristic x can be seen as a basic sensory input such that categories are formed based on a relationships between the characteristics of the cases (e.g. like a metric on the characteristics space or feature overlaps, etc.). In this way, for instance a very specific procedure to evoke categories might solely activates categories that contain a case that coincides (with regard to the characteristic part) with the current problem, i.e. $\tilde{C}_x$ is the list of categories, that contain at least one case with characteristic $x \in X$, i.e. let $A_l := \{x \in X | \exists c = (x, r) \in C \text{ s. th. } c \in \tilde{C}_l\}$ for all $l \leq L$, then

$$\tilde{C}_x := (1_{A_1}(x)\tilde{C}_1, 1_{A_2}(x)\tilde{C}_2, ..., 1_{A_L}(x)\tilde{C}_L) \in C^*.$$

Such an procedure would imply that cases with different characteristics than x are only activated if they are contained in categories that include member cases with characteristic (problem) x.[15] However, our definition is general and thus may not necessary rely on the

[13] Some research on this issue is mentioned in the introduction. Note that under automaticity subjects are often not even aware of this process.
[14] However, an outcome dependent categorization is in principle also possible.
[15] Such a categorization is only useful, if cases with different characteristics into the same categories.

x characteristics specifically, but may take into account the categories that are activated by any characteristics that are close or related to x. Furthermore another widely accepted procedure depends on the closest "distance" (with respect to similar, salient, related, familiar) to the prototypical element of a categories (e.g. Rosch and Lloyd (1978)), which then trigger their corresponding category(ies) members.

Properties on problem activated categories

Now we define a consistent transfer to activations of the categories in a database. Based on a natural categorization structure $\tilde{C} = \{\tilde{C}_1, \tilde{C}_2, ..., \tilde{C}_L\}$, we will define a function $\tilde{C}: X \times C^* \to C^*$ that determines the single pieces of information in a database $D \in C^*$ that belong to categories that are activated by a specific problem $x \in X$.

Thereby, our main assumption concerns a minimal amount of distinct cases in a database that initiates categorical processing of information in agents' minds. For a less diverse or complex database (i.e. $div(D) \leq k$) an agent's brain does not start to simplify and reduce the set of information by categorizing the information, but will just process, inspect and take into account all single pieces of information directly. However, for more diverse or complex databases agent's mind will initiate a rough (problem evoked) classification of the database in accordance with the natural categorization $\tilde{C}$ and then consider in detail only the information in her database that are also contained in categories activated by the current problem.

Definition 5.2 ***Induced minimal categorization***
Let $\tilde{C}$ be a categorization structure on C and $k \in \mathbb{N}$ with $k \geq 3$. A database $\tilde{C}(x, D)$ results from a categorization function $\tilde{C}: X \times C^* \to C^*$[16] *that categorizes each database $D \in C^*$ according to the categories in $\tilde{C}_x$ that are evoked by problem $x \in X$. We call $\tilde{C}(x, D)$ the x-evoked/activated categorized database D (for cognitive ability level k) and define it for all $D \in C^T$ for all $T \in \mathbb{N}$ by*

$$\tilde{C}(x, D) := \begin{cases} D \cap (c^T)_{c \in C} & \text{for } div(D) \leq k \\ D \cap (\tilde{C}_x)^T & \text{for } div(D) > k \end{cases}$$

In the following, we will fix $k = 3$ without loss of generality. Obviously $\tilde{C}(x, D) \subseteq D$.

Example:
In order to clarify the definition of the categorization function for a database $D \in C^T$ such that $div(D) > 3$, i.e. $\tilde{C}(x, D) := (\tilde{C}_x)^T \cap D$, consider the following situation. Let $D = (c_1^2, c_2, c_3^4, c_6, c_7^3) \in C^{11}$, i.e. $div(D) = 5$, and a categorization structure induced by problem $x \in X$ $\tilde{C}_x = (\tilde{C}_1, \tilde{C}_2)$, where $\tilde{C}_1 = (c_1, c_2)$ and $\tilde{C}_2 = (c_2, c_3, c_4, c_7)$. Then $(\tilde{C}_x)^{11} \cap D = (c_1, c_2)^{11} \circ (c_2, c_3, c_4, c_7)^{11} \cap D = (c_1^2, c_2, c_3^4, c_7^3)$, which is normatively and descriptively appealing for such a categorized database.

Remark 5.1
(i) Since $\tilde{C}_x$ is insensitive to repetitions of cases, we need to ensure that repeated observa-

[16] With slight abuse of notation we name the function and the natural categorization identical to emphasize that for each fixed categorization structure a corresponding function can be defined, i.e. that the function relies on this fixed categorization structure like a parameter.

tions in $D \in C^T$ are captured, which is guaranteed by introducing the T-replicated $\tilde{C}_x$.[17]
(ii) An immediate consequence of the minimal induced categorization property is that for database D such that $div(D) \leq 3$, we have no categorization and the framework (and later defined axioms) coincides with the framework of BGSS. Thus, it enables us to mirror their proof for these kind of databases.

Apart from the minimality condition for an activated categorization of databases, the definition makes implicitly three additional important assumptions.

First, a very important ingredient of this definition is that the evoked categories $\tilde{C}_x$ are totally unrelated to the database under consideration. Only the underlying problem activates the relevant categories. In accordance with these categories the actually available information in the database is intersected. One may argue that a database itself determines which categories are evoked (or even formed), since the database might provide some intuition and motivation how to categorize it. However, our intuition runs solely through a new problem x that activates the relevant categories in the subconsciously pre-existing natural categorization $\tilde{C}$. In this sense the induced categorization of the database occurs not directly on the level of the database.

Second, the ordering of the database D does not affect the resulting categorized database. Any reordered database $\pi(D)$ of the database D results identically activated categorized information, i.e. the content of the categorized database $\tilde{C}(x, D)$ and $\tilde{C}(x, \pi(D))$ coincides. In this sense, the definition induces some categorization invariance that is driven by our assumption that the categorization is evoked by the underlying problem x and not by the database. This precludes that a categorization of a database is affected by order effects.[18] However, since the x-evoked categories are activated independent from any database in our approach,it is quite natural that their intersections with any reordered database result in the same content.

Finally however, even though the content of differently ordered database are identical after the categorization, the order of its evoked content (cases) in the categorized databases may still be different, i.e. $\tilde{C}(x, D)$ and $\tilde{C}(x, \pi(D))$ may consist of the same content but differently ordered. However, the definition precludes this difference by assuming a specific ordering for all categorized databases according to the order on the set of basic cases (induced by the definition of $\cap$ for databases and the order on C). Yet, the reason for this assumption is not that we want to restrict the categorization process on databases in this way, but it is rather an anticipation of a property or axiom we would enforce for the subsequent belief formation. In the manner of BGSS, an Invariance Axiom on beliefs would say that the belief induced by a database is determined only by its content and not its order of information in the database, i.e. the beliefs induced by any reordering of the same content coincides. Thus, the assumption of a specific (seemingly restrictive) order of the categorized databases is quite innocent and harmless in combination with an Invariance Axiom on the belief stage, since then the order of any categorized database is immaterial for a belief.

[17] We use the total length of the database just for simplicity. One might also take the maximal amount of a case appears in D.

[18] For instance a first impression effect might might induce a bias for the category that the first case in the database is most related to.

5.2 Induced category based belief

A category based belief is composed of a usual belief $P : X \times C^* \to \Delta(R)$ and a previous problem evoked categorization of the underlying information $\tilde{C} : X \times C^* \to C^*$ (i.e. $\tilde{C}(x, D) \subseteq D$), such that $(P \circ \tilde{C}) : X \times C^* \to \Delta(R)$, i.e. $(P \circ \tilde{C})(x, D) = P(x, \tilde{C}(x, D))$. Faced with a new problem $x \in X$, the agent's brain activates or evokes some appropriate categories for this problem according to the natural categorization structure $\tilde{C}$ and forms the belief based only on those pieces of information in the activated categories that are actually available in her database, i.e. on sub-database $\tilde{C}(x, D) \subseteq D$.

In the following we will fix a problem $x \in X$ and write for convenience $\tilde{C}(x, D) = \tilde{C}(D)$ and for $(P \circ \tilde{C})(x, D) = (P \circ \tilde{C})(D)$ when no confusion arises.

5.3 Axioms on the level of belief formation

Categorized Invariance Axiom (already implied)
For all $D \in C^*$ and all permutations π on D, i.e. $D = (c_1, ..., c_T)$, then $\pi(D) = (c_{\pi(1)}, ..., c_{\pi(T)})$ the following holds:

$$(P \circ \tilde{C})(D) = (P \circ \tilde{C})(\pi(D))$$

The axiom basically says that the order or sequence of appearance of the cases in D is immaterial for the induced category based belief, only the content matters. Thus, the axiom is directly implied by our definition of a problem evoked categorization of databases, since we discussed already that for any two databases containing the same content, their induced categorization coincide, i.e. $\tilde{C}(D) = \tilde{C}(\pi(D))$ for all databases $D \in C^*$ and reordered database $\pi(D)$. In this sense the categorized Invariance Axiom is superfluous and indirectly substituted by the definition of a categorized database.[19]

Per se the invariance property does not allow for different impacts if a case appears earlier or later in a database. However the order in which information is provided or obtained can influence the judgment strongly and may carry information by itself. One way to cope with these order effects is to describe the cases informative enough. E.g. if one wants to capture the position or time of occurrence of a case in the categorized database, one could implement this information into the description of the cases itself. Put differently, if one challenges the consequences of an invariance property, then there must be some criteria which distinguishes the cases and paying attention explicitly to this difference in the description of the case may lead the agent to reconcile with such an invariance.

Category based Concatenation Axiom
There exists some $\lambda \in [0, 1]$, such that for $\tilde{C}(D \circ E) \neq \emptyset$

$$(P \circ \tilde{C})(D \circ E) = \lambda (P \circ \tilde{C})(D) + (1 - \lambda)(P \circ \tilde{C})(E),$$

where $\lambda = 0$ if and only if $\tilde{C}(D) = \emptyset$.

[19]Technically speaking, there is no difference between restricting the categorization process to specific orders or allowing for different orderings and requiring an Invariance Axiom for beliefs.

In the following we will call the database which emerges from concatenation of other databases as the **combined or concatenated** database, whereas the databases used for the concatenation will be called **combining or concatenating** databases.

The category based Concatenation Axiom states that a category based belief induced by a concatenated database is a weighted average of the category based beliefs induced by its respective combining databases. The axiom captures the idea that a belief based on the combination of two databases can not lie outside the interval spanned by the beliefs induced by each combining database separately. Intuitively it can be interpreted in the following way (stated from an exclusion point of view): if the information in any database induces an agent's belief not to exclude an outcome r, then the outcome r cannot be excluded by the belief induced by the combination of all these databases.[20]

However, in order to sustain the normative appealing interpretation of averaging (category based) beliefs, the categorized concatenation of two databases must coincide with the concatenation of these two categorized databases, i.e. the union of the elements surviving the categorization process for each single database should not differ from the elements surviving the categorization of the database generated by the combination of the two. This would ensure that a category based belief induced by the concatenated database relies on information that is also employed in the category based beliefs induced by the single concatenating databases. However this is directly achieved by the definition of a problem evoked categorization of a database, i.e. for $D \in C^T$ and $E \in C^L$

$$\tilde{C}(x, D \circ E) = (\tilde{C}_x)^{T+L} \cap (D \circ E) = ((\tilde{C}_x)^T \cap D) \circ ((\tilde{C}_x)^L \cap E) = \tilde{C}(x, D) \circ \tilde{C}(x, E)$$

The category based Concatenation Axiom assumes $\lambda = 0$ for databases such that $\tilde{C}(D) = \emptyset$. Of course $\lambda \neq 0$ would result in inconsistencies, since then $(P \circ \tilde{C})(D \circ E) = \lambda P(\emptyset) + (1 - \lambda)(P \circ \tilde{C})(E)$, which implicitly states that the category based beliefs induced by $\tilde{C}(D \circ E)$ and $\tilde{C}(E)$ would differ, even though the categorized databases coincide.

Collinearity Axiom
No three of $((P \circ \tilde{C})(c))_{c \in C}$ such that $\tilde{C}(c) \neq \emptyset$ are collinear.

Technically speaking this axiom allows to derive an unique similarity function (in combination with the other axioms), but it has also some reasonable intuition. Roughly it states that a (non trivial) estimate based on a case is never equivalent to the combined (non-trivial) estimates based on two other cases. Hence, a case is always informative in the sense that no combination of two other cases can deliver the same estimation and would make this case redundant. By non trivial we mean that the case is activated, since a not activated case could only contribute a trivial (uninformed) uniform-like estimate.

[20]Of course the axiom is stronger in the sense, that it not only requires that the probability of such an r is positive, but it should lie between the minimal and maximal assigned probabilities induced by the combining (filtered) databases.

5.4 Representation Theorem of category based belief formation

Theorem 5.1

Let there be a function $(P\circ\tilde{C}) : C^* \to \Delta(R)$*, where* $P : C^* \to \Delta(R)$ *and* $\tilde{C} = \{\tilde{C}_1, ..., \tilde{C}_L\}$ *a categorization structure on* C *with corresponding induced minimal categorization function* $\tilde{C} : C^* \to C^*$*, i.e. for each* $D \in C^*$ *a categorized database* $\tilde{C}(D) \subseteq C^*$ *is given. Let* $(P \circ \tilde{C}) : C^* \to \Delta(R)$ *satisfy the Collinearity Axiom.*
Then the following are equivalent:
(i) The function $(P \circ \tilde{C})$ *satisfies the category based Concatenation Axiom*
(ii) There exists for each $c \in C$ *a unique* $P^c \in \Delta(R)$*, and a unique strictly positive -up to multiplication by a strictly positive number- function* $s : C \to \mathbb{R}_+$*, such that for all* $D \in C^*$ *with* $\tilde{C}(D) \neq \emptyset$

$$(P \circ \tilde{C})(D) \quad = \quad \frac{\sum_{c\in\tilde{C}(D)} s(c)P^c}{\sum_{c\in\tilde{C}(D)} s(c)}.$$

Rough sketch of the proof:

The necessity part is straightforward calculation. The sufficiency part follows the rough structure of the proof of BGSS and Bleile (2014b), but differs in the crucial arguments. The idea is to transform the framework from the space of databases to the space of frequency vectors that is structural more tractable, i.e. the filtered belief based on databases $(P\circ\tilde{C})(D) = \frac{\sum_{c\in\tilde{C}(D)} s(c)P^c}{\sum_{c\in\tilde{C}(D)} s(c)(c)}$ for $D \in C^*$ translates to frequency vectors f_D by $(P\circ\tilde{C})(f_D) = \frac{\sum_{j\leq m} s_j\tilde{C}_j(f_D)P^j}{\sum_{j\leq m} s_j\tilde{C}_j(f_D)}$. In order to show that this is viable we exploit the structure of the categorization procedure and the categorized Concatenation Axiom.
The essential part of the proof is to derive the similarity weights $(s_i)_{i\leq m}$. This will be shown inductively over $|C| = m$ and $div(f_E) \leq m$.

Step 1: Base case for the induction, i.e. $|C| = m = 3$. Since $\tilde{C}(f) = f$ for all f such that $div(f) \leq 3$, we are exactly in the BGSS framework, which directly deliver the result for these kind of frequency vectors.

Step 2: $|C| = m > 3$ and $div(f_E) \leq m$.
As in BGSS, we can show (using $\tilde{C}(f) = f$ for all f such that $div(f) \leq 3$) that the similarity weights derived in Step 1 are independent of the triplet $\{i, j, k\}$ for any set of basic cases $C = \{c^i, c^j, c^k\}$ and thus we can define for all $f \in \Delta(C)$

$$(P \circ \tilde{C})_s(f) := \frac{\sum_{j\leq m} s_j\tilde{C}_j(f)P^j}{\sum_{j\leq m} s_j\tilde{C}_j(f)}.$$

The aim is to show $(P \circ \tilde{C})_s(f) = (P \circ \tilde{C})(f)$ for all $f \in \Delta(C)$ via induction over m and using Step 1 ($m = 3$) as base case.
Let $f = \alpha q^j + (1 - \alpha)f(j)$ (*) (for some $\alpha \in (0, 1)$) where $f(j)$ denotes the point in $conv(\{(q^l)_{l\in\{1,...,m\}\setminus j}\})$ that is on the line through f and q^j, as in BGSS.
(i) If there exists a $j \leq m$ such that $q^j \notin \tilde{C}(f)$, then the decomposition (*) and the category based Concatenation Axiom (and induction assumption) delivers the claim.

(ii) If $q^j \in \tilde{C}(f)$ for all $j \leq m$, then there are m many q^j such that $\tilde{C}(q^j) \neq 0^m$. Again, the category based Concatenation Axiom applied to the m many decompositions yields that $(P \circ \tilde{C})(f)$ lies in the interior of the intervals spanned by $((P \circ \tilde{C})(q^j), (P \circ \tilde{C})(f(j)))$ for all $j \leq m$. Combined with the Collinearity Axiom this delivers a unique intersection of these lines in $(P \circ \tilde{C})(f)$ and $(P \circ \tilde{C})_s(f)$, since the elements determining the lines satisfy already the claim by the induction assumption.

Interpretation of Theorem

A category based belief formation can be interpreted as a two stage process in which in an initial step the rough categorized information is activated by the current problem and in a subsequent step the information contained in these activated categories are processed and evaluated in detail for the belief formation.
A category based belief follows exactly the experimental evidence in psychology in which individuals focus on the category(ies) that a problem belongs to or are most relevant and fitting (Murphy and Ross (1994)) and all other categories are immaterial, not retrieved and excluded. This implies that an agent does not need to retrieve or consider all potentially memorized or all past cases (as in BGSS, EG, Bleile (2014a)). In this way, such a procedure may reduce enormously the cognitive effort, since only a subset of past cases is processed in detail and all pieces of information that are members of irrelevant categories are not even needed to be retrieved.

A category based belief is not based on estimations that are associated with entire categories, but it relies on all estimates induced by the single cases in the categories activated by the problem. This is an important distinction to the axiomatized prototype based belief in the next section and reflects the disagreement in the categorization literature on how categories are actually represented. One stream of literature argues for a representation through all its members (Kruschke (1992), Medin and Schaffer (1978)), whereas another branch reasons in favor of an abstracted summary in terms of a prototype representation (e.g. Rosch and Mervis (1975) and references in Murphy (1994)).[21]

Since a category based belief relies only on the information of some activated categories, it can be interpreted as a limited attention model as in Bleile (2014b), where a filtered belief is formed based only on some parts of the potentially available information (so called consideration set that survives a screening/filtering stage).[22] For a category based belief, the information in the evoked categories can be identified as the information that is contained in such a consideration set in the limited attention model. From this point of view, the filtering runs on the category level and the categories are roughly screened with the purpose to "determine" whether they are appropriate or not for the current problem. The "surviving" categories are examined in full detail in the further belief formation. Nevertheless, the category based and filtered belief are interpretational very similar, however from a structural, technical and axiomatic point of view they differ significantly. The cat-

[21] In general, there is tradeoff between informativeness and economy involved that might be better balanced by a more intermediary representation as offered in VAM models (Vanpaemel and Storms (2008)), which allows for varying abstraction levels. It seems that people shift from using a prototype representation early in training to using an exemplar representation late in training.

[22] A filtered belief $(P \circ \Gamma^E)(x, D)$ that is based also only on some parts $\Gamma^E(D) \subseteq D$ of the information captured in D, since it is filtered in some way, i.e. $(P \circ \Gamma) : X \times C^* \times C^* \to \Delta(R)$, but requires some different setup.

egory based belief formation can also be interpreted as an adaption of the choice model of Manzini and Mariotti (2012) ("Categorize then choose") to belief formation.

6 Axiomatization of prototype based belief formation

6.1 Specific properties of the categorization procedure

For a prototype based approach we will define a specific categorization of databases in accordance with a natural categorization structure, but independent of the current problem of categorizing information. In this regard it differs from the first procedure.

Natural categorization structure

Let there be given a natural (evolutionary optimal) categorization structure $\tilde{C}$ as discussed in Section 4. We slightly restrict the natural categorization structure such that it satisfies some additional structural properties regarding the content of the categories.

Definition 6.1
For a set of basic cases C we call $\tilde{C} = \{\tilde{C}_1, ..., \tilde{C}_L\}$ for some $L \in \mathbb{N}$ a categorization structure, if
(i) $\tilde{C}_k \cap \tilde{C}_l = \emptyset$ and
(ii) $C = \circ_{j \leq L} \tilde{C}_l$ and
(iii) for all $l < L$ $|\tilde{C}_l| \geq 2$ and $|\tilde{C}_L| \geq 1$

In contrast to the unrestricted natural categorization structure in the section before, now the categories are explicitly defined to be disjoint and should contain sufficiently many elements. Disjointness is natural when a category is identified by a (set of) property or attribute. An object does either possess a property or it does not, which implies the disjointness. We stick to disjoint categories mainly for reasons of technical and notational simplicity, but it could be generalized.
The assumption that almost all categories contain at least two members captures the motivation to deal with "real" categories.[23] Basically, a "real" category in our definition exists whenever at least two or more heterogenous or distinguishable cases can be gathered in the same category according to some common criteria. Categorization is only meaningful if some pieces of information can be classified into a common genuine category. We rule out a degenerate (trivial) singleton-categorization, in which all cases get their own category. Furthermore, the optimal categorization literature (Fryer and Jackson (2008), Mohlin (2014), Peski (2011)) supports our defining properties (under some mild conditions). It shows that there exist many more cases than categories and that an optimal categorization results in few and relatively coarse categories, which implies that "optimal" categories should contain many members, i.e. $|\tilde{C}_l| \geq 2$.

The underlying reason originates from a tradeoff between benefits and disadvantages of a fine or coarse categorization.[24] A finer categorization implies more categories that contain

[23] Manzini and Mariotti (2012) consider a similar requirement for their 2nd version of categorization.
[24] This is closely relate to the problem of over-fitting in statistics, where a too close/precise fit of limited observations- i.e. using high-dimensional models- comes with risk of loosing the predictive power.

less but more homogenous and similar members, but results in a decreasing robustness or reliability of a prediction based on these (less precise or noisy) categories with less many observations. Another compromise concerns the increasing challenge in searching and identifying the "correct" category(ies) for new objects for finer (and more narrow) categorization (e.g. Medin (1983), Jones (1983)). Thus an agent might prefer to categorize more coarsely into larger categories, which is well known and discussed in the psychology literature and referred to as basic level categories. Basic level categories are neither the most general nor the most detailed categories.

A categorization based solely on characteristics satisfies our definition under the condition that the L categories are not empty, since if a case $c = (x, r) \in \tilde{C}_l$, then the cases $c = (x, r_i)$ are in the category $\tilde{C}_l$ for all $i \leq n$.

The assumption that there exists at most one singleton category reflects the intuition of a category that might collect the cases that are "uncategorizable" into "real" categories. This is supported by an implication of optimal categorizations (e.g. Fryer and Jackson (2008)) which shows that experiences and objects in databases that are not "easy" to categorize (i.e. tend to form a singleton category) are more coarsely categorized and more often lumped together (i.e. gathered in the category of uncategorizable elements $\tilde{C}_{L+1}$).

Specific database induced categorization

Based on a natural categorization structure, we define a specific categorization procedure for given databases that transmits the idea that there is one category that contains the "uncategorizable" elements. Further it assumes -similar to the problem evoked categorization of databases in Section 5- that a minimum amount of complexity of the database is required in order to initiate categorical thinking and processing of information. For less diverse databases categorization is not necessary and the agent considers just all cases in detail.

Definition 6.2
Let $\tilde{C} = \{\tilde{C}_1, .., \tilde{C}_L\}$ be a categorization structure on C. For all $E \in C^*$ a categorization of E or E-categorization structure $\tilde{C}^E = \{\tilde{C}_1^E, ..., \tilde{C}_{L+1}^E\}$ is given in the following way:
(i) If $div(E) < 5$, i.e. $E = (c_i^r, c_j^s, c_k^t, c_n^u)$ for distinct $i, j, k, n \leq m$ and $r, s, t, u \in \mathbb{N}_0$, then $\tilde{C}^E = \{\tilde{C}_1^E = \{c_i\}, \tilde{C}_2^E = \{c_j\}, \tilde{C}_3^E = \{c_k\}, \tilde{C}_4^E = \{c_n\}, \tilde{C}_5^E = \emptyset, ..., \tilde{C}_{L+1}^E = \emptyset\}$
Basically $\tilde{C}^E = \{\{c\}_{c \in E}\}$
(ii) If $div(E) \geq 5$, then $\tilde{C}^E := \{\tilde{C}_1^E, ..., \tilde{C}_{L+1}^E\}$, where $\tilde{C}_l^E$ for $l \leq L$ is defined as follows:

$$\tilde{C}_l^E = \begin{cases} \tilde{C}_l \cap E & \text{if } div(\tilde{C}_l \cap E) \geq 2 \\ \emptyset & \text{if } div(\tilde{C}_l \cap E) \leq 1 \end{cases}$$

$$\tilde{C}_{L+1}^E = \bigcup_{\{l \leq L \,|\, |\tilde{C}_l \cap E| = 1\}} \tilde{C}_l \cap E$$

Note, that by definition $\tilde{C}_l^E$ (for all $l \leq L+1$) does not contain repetitions of cases.
The difference in processing information depending on $div(D) \lesseqgtr 5$ captures the motivation to have at least three meaningful categories, in the sense that at least two "real" non singleton categories exist, which requires $div(D) \geq 5$.

Remark 6.1

The definition of a categorization of a database implies some sort of categorization invariance, i.e. categorizations are independent of the order, number and frequency of cases in a database. More precisely,

(i) $\tilde{C}^E = \tilde{C}^D$ *for all* $D, E \in C^*$ *containing the same cases, i.e. for all* $c \in C$ $f_D(c) > 0$ *if and only if* $f_E(c) > 0$*. In particular* $\tilde{C}^E = \tilde{C}^{E^Z}$ *and*

(ii) $\tilde{C}^E = \tilde{C}^{\pi(E)}$ *holds for all re-orderings* $\pi(E)$ *on E.*

Basically, we require that a category (or later its prototype) is not affected by repetitions of already observed information, indirectly saying that categories are characterized by single observations of different cases and not influenced by their frequencies. Interestingly, an optimal categorization procedure (a la Jackson and Fryer (2008)) results as well in categories that remain unchanged when experiences is simply replicated (and also their prototypes), i.e. $\tilde{C}_l^{D^Z} = \tilde{C}_l^D$ for some $Z \in \mathbb{N}$.[25]

The categories of an E-categorization structure (that an agent has in mind) can be evoked or activated by cases that are also contained in another (simultaneously) available (somehow related) database D.

Definition 6.3

Let $D, E \in C^*$*, such that* $f_D(c_i) \geq 0$ *only if* $f_E(c_i) > 0$ *for all* $i \leq m$*. Then, the E-categories evoked by D result from an D-induced E-categorization function* $\tilde{C} : C^* \times C^* \to P(C^*)$

$$\tilde{C}(D, E) := \tilde{C}^E(D) = (\tilde{C}_1^E(D), .., \tilde{C}_{L+1}^E(D)), \quad \textit{where for all } l \leq L+1$$

$$\tilde{C}_l^E(D) \quad := \quad \begin{cases} \tilde{C}_l^E & \textit{if } D \cap \tilde{C}_l^E \neq \emptyset \\ \emptyset & \textit{otherwise} \end{cases}$$

The definition is basically some consistency condition (note that $\tilde{C}_l^E(E) = \tilde{C}_l^E$), but one can interpret it as well in the following way. An agent having already categorized a database E and is "simultaneously" faced with processing the database D (consisting only of already categorized cases in E) will not forget her already "internalized" E-categorization structure. In particular, the cases in D activate some categories in the richer E-categorization structure $\tilde{C}^E$ and it might only happen that some of these categories are not activated by cases in D, i.e. if for a $l \leq L+1$ and all $c \in D$ $c \notin \tilde{C}_l^E$, then $\tilde{C}_l^E$ is not evoked by D. However this interpretation does only apply for "simultaneously" available and actively categorized databases. This is distinct to a This interpretation doe not apply for the natural categorization structure that is subconsciously (evolutionary and automatically) anchored in the brain. and thus not actively formed in a process.

An important implication of the definition is that for any reordering $\pi(D)$ of the database D, we have the same list of evoked categories in terms of content as well as in terms of the order, i.e. $\tilde{C}^E(D) = \tilde{C}^E(\pi(D))$. In combination with the Remark 6.1, this implies a categorization invariance, i.e. for any $(D, E) \in C^* \times C^*$ such that $f_D(c) \geq 0$ if $f_E(c) > 0$

[25] However, the optimal categorization of Fryer and Jackson is sensitive to additional already known information, if it concerns only single pieces of information, i.e. (i). Increasing the size of only a single group of cases may lead to a shift in the categorization.

and appropriate permutations π, π'

$$\tilde{C}^E(D) = \tilde{C}^{E^L}(D^Z) = \tilde{C}^{\pi(E)}(\pi'(D)) \tag{1}$$

This property of a database evoked categorization of information is restrictive, since the order of cases in D might affect the order in which the E-categories are activated.[26] However, similar as in Section 5, this specific assumption is not a property we want to enforce explicitly, but it is an anticipation of an Invariance Axiom on the belief level, we would enforced if the evoked categorization structure would be order sensitive.

Admissability of database based categorization

We defined an environment in which pairs of (somehow nested) databases $(D, E) \in C^* \times C^*$ affect the belief formation. The richer database E induces some categorization, where the cases of D activate the E-categories for the actual process of belief formation. In such a framework not all potential combinations of databases are plausible and meaningful for a belief formation based on categorized information. Our admissibility condition specifies the circumstances under which categorization of information is a normatively and descriptively reasonable.

As discussed already, a categorization heuristics is useful for sufficiently complex and diverse databases. However a sufficiently diverse database is not directly complex, e.g. if it is classified into a single (or very few) large category (ies) or into almost singleton (very fine) categories. Basically, only if sufficiently many meaningful categories are evoked an agent starts to think and process information categorical and feels confident in relying on (summarized) information on the category level. For databases that involve only very few activated databases an agent may not want to rely on only coarse (imprecise and noisy) summaries of these few categories, but might go through the information case by case in order to be sufficiently informed.[27] Thus, in such a situation a categorization of information does not offer some advantage to an approach of just taking into account all single pieces of information directly.

Our admissibility condition (i.e. (ii) and (iii)) restricts the pairs of databases for which an agent starts the categorical thinking an processing. It requires that a minimum number of "real" categories of $\tilde{C}^E$ (namely three)are in some sense activated by a database D and considered for its evaluation.

Definition 6.4
The admissible pairs $(D, E) \in C^* \times C^*$ is given by the set A as follows

$$\begin{aligned} A := \{(D, E) \in C^* \times C^* | \quad & (i) \quad f_D(c) > 0 \text{ then } f_E(c) > 0 \text{ for all } c \in C \\ & (ii) \quad \text{if } div(D) = 2 \text{ then there must exist } c \in E \backslash D \\ & \qquad \text{such that } |\tilde{C}^E(D \circ c)| = 3 \\ & (iii) \quad \text{if } div(D) \geq 3 \text{ then } |\tilde{C}^E(D)| \geq 3\} \end{aligned}$$

[26] With additional effort on notation and definitions we could take care of orders.

[27] As in the setup of a category based belief or the original belief formation without a categorization as taken in BGSS, EG, Bleile (2014a)

Note that $(D, E) \in A$ if and only if $(D^Z, E^L) \in A$.
A necessary condition for a "real" categorization is $div(E) \geq 5$. For less diverse E each contained case is interpreted as singleton category and all conditions in A are naturally satisfied. Thus, we need to discuss the admissibility conditions only for more diverse databases that induce non trivial singleton categories.

For condition (ii) and (iii) to be satisfied we need $|\tilde{C}^E| \geq 3$, which captures our motivation to employ only sufficiently "rich" categorizations. Consequently, those pairs (D, E) are ruled out such that $|\tilde{C}^E| < 3$. Part (iii) captures explicitly the intuition of "satisfactorily" many activated E-categories, by enforcing that there must exist at least three different cases in D that evoke three different E-categories.

For database D such that $div(D) = 2$ part (ii) requires that both contained different cases need to belong to different categories according to the E-categorization structure. The underlying idea is that D activates two different E-categories and thus triggers (makes aware) some categorical thinking and processing. However, if the database D evokes just one category $\tilde{C}_l^E$ in the E-categorization, then an agent might not initiate any categorical thinking and processing at all or is just not aware of different categories, but might rely on each single case directly. This is exactly ruled out by condition (ii).[28]

The admissibility condition is justifiable in general, but our initial motivation originates in the most interesting situation for $D = E$. The condition (i) is directly met. A pair (D, D) with $div(D) = 2$ is not admissible. This matches our desire that no categorization is induced for databases that has no sufficiently complex categorization - i.e. not at least three categories- such that a categorization procedure offers (summarized) information on the category level for acceptable many categories. For more diverse databases, $div(D) \geq 3$, the admissibility just requires that the D-categorization structure consists of at least three different categories, as we desired.

6.2 Induced prototype based belief

An agent forms a prototype based belief based on her available admissible pair of information $(D, E) \in C^* \times C^*$ in the following way. Based on a natural categorization structure $\tilde{C}$, a categorization of a database E result in the categorization $\tilde{C}^E$. An agent evaluates the simultaneously available database D by exploiting the categorized information $\tilde{C}^E$ contained in the richer database E that is activated by cases in D.
Thus, a prototype based belief relies on categories in E that are evoked/activated by the D induced categorization function $\tilde{C} : C^* \times C^* \to C^*$, (i.e. $\tilde{C}(D, E) = \tilde{C}^E(D)$) such that $(P \circ \tilde{C}) : X \times C^* \times C^* \to \Delta(R)$, i.e. $(P \circ \tilde{C})(x, D, E) = P(x, \tilde{C}(D, E)) = P(x, \tilde{C}^E(D))$.[29]
$(P \circ \tilde{C})$ is a belief induced by the categories in the richest database available (i.e. E) that are evoked/activated by single pieces of information in the database under consideration(i.e. D). The most intuitive situation is where $E = D$, in which all D-induced categories are employed.

[28]Note, that for all $D \subseteq E \in C^*$ with $div(D) = 1$ (D, E) is admissible, in particular for any $E \in C^*$ and all $c \in E$ $(c, E) \in A$. This requires an exception of the given interpretation, since it is obvious that only one E-category can be activated. However, this is driven by the situation $D = E = (c^T)$, which implies a belief only based on information (and only category member) c, which is a acceptable in the light of the desired representation in Theorem 6.1.
[29]I.e. $P : X \times P(C^*) \to \Delta(R)$ relies only on categories.

Throughout the paper the problem x is fixed, therefore x is often suppressed in the following, i.e. $(P \circ \tilde{C})(x, D, E) = (P \circ \tilde{C})(D, E)$.

6.3 Axioms on the level of belief formation

As already mentioned above, our (restrictive) assumptions on the categorization procedure (see Remark and 6.1 equation (1)) replaces the otherwise imposed Invariance Axiom.

Implied Invariance Axiom
For all admissible pairs $(D, E) \in A$ and $D \in C^T$ and all permutations π on T for all T, we have

$$(P \circ \tilde{C})(D, E) = (P \circ \tilde{C})(\pi(D), E)$$

The Invariance Axiom says that the order of the information in the database is immaterial for the induced belief. Only their content is important. For a discussion of the axiom see Section 4.2.

Remark 6.2
Note, that in particular $(P \circ \tilde{C})(D, E) = (P \circ \tilde{C})(\pi(D), \pi'(E))$ *by Remark 6.1).*

Thus the order invariance accounts for both databases, which is important for the proof.

Prototype based Concatenation Axiom:
Let $\tilde{C}$ ba a categorization structure on C. For all $D, E, F \in C^*$ such that $(D, F) \in A$ and $(E, F) \in A$ are admissible pairs, then there exist $\lambda \in (0, 1)$ such that:

$$(P \circ \tilde{C})(D \circ E, F) = \lambda(P \circ D)(D, F) + (1 - \lambda)(P \circ \tilde{C})(E, F)$$

The interpretation of the axiom is similar to the Concatenation Axiom in BGSS and as in Section 4.2. In order to keep the normatively desirable spirit of averaging, we need to ensure that the information employed in the belief formation for the concatenation $D \circ E$ is meaningfully related to the single databases D and E. Intuitively, the categories evoked by $(D \circ E)$ need to be covered by the categories evoked by either D or E. This can only hold in general, if a common categorization structure underlies all involved activation processes, i.e. for a common $\tilde{C}^F$. A determination of a belief as an average of two other beliefs would be hard to justify if the underlying beliefs rely on different categorizations. If so, there might be very different categories involved in the different beliefs that would prevent an easy averaging, since no common evaluation basis exists. Of course, to be able to activate some categories from this common categorization structure, all observed cases in D and E need to be categorizable with regard to this common basis. Thus, the F-categorization ought to cover at least the available information in $D \circ E$, i.e. $F \supseteq D \circ E$. Moreover, having this categorization structure in mind, it appears reasonable to employ it in the belief formation process and do not shift to another less rich categorization, e.g. like moving to a E-categorization for database E, i.e. $\tilde{C}^E(E)$. Thus, the assumed structure ensures that $\tilde{C}^F(D \circ E) = \tilde{C}^F(D) \cup \tilde{C}^F(E)$ and therefore the induced beliefs rely on the same F-categories and are only distinct in the way which categories their contained cases evoke. The prefix "prototype based" Concatenation Axiom or belief will become clear in Theorem 6.1 and would allow for an interpretation of the axiom in terms of category related prototypes.

Identity Axiom

Let $(D_1, E), (D_2, E) \in A$ and related categorization structure $\tilde{C}^E = \{\tilde{C}_1^E, .., \tilde{C}_{L+1}^E\}$. For all D_1, D_2 such that $\tilde{C}^E(D_1) = \tilde{C}^E(D_2) = \tilde{C}_l^E$ for some $l \leq L+1$, then $(P \circ \tilde{C})(D_1, E) = (P \circ \tilde{C})(D_2, E)$.

The axiom says that the prototype based beliefs induced by databases (given an E-categorization structure) coincide, if the databases evoke only one identical E-category. In this situation, the specific content of the information is immaterial for the induced belief, since only the activated category is relevant.

Collinearity Axiom

For all databases $D \in C^*$, no three distinct vectors of $((P \circ \tilde{C})(c, D))_{c \in D}$ are collinear.

The interpretation is the same as in Section 4.2 or BGSS.
The only difference is the requirement of distinctiveness, since $(P \circ \tilde{C})(c, D)$ is identical for cases in D that are contained in (and activate) the same category. Basically, no three estimates based on different categories are collinear.

6.4 Representation Theorem of prototype based belief formation

Theorem 6.1

Let $(P \circ \tilde{C})$ be a function $(P \circ \tilde{C}) : C^* \times C^* \to \Delta(R)$, where $\tilde{C} : C^* \times C^* \to P(C^*)$ is an induced categorization function with underlying $\tilde{C}^E = \{\tilde{C}_1^E, ..., \tilde{C}_{L+1}^E\}$ for all $E \in C^*$. Let the prototype based belief $(P \circ \tilde{C})$ satisfy the Collinearity Axiom.
Then the following are equivalent
(i) $(P \circ \tilde{C})$ satisfies the prototype based Concatenation and Identity Axiom
(ii) For all $E \in C^*$ and any $l \leq L+1$ such that $\tilde{C}_l^E \neq \emptyset$ there exists a unique $P^{\tilde{C}_l^E} \in \Delta(R)$ and a unique -up to multiplication by a strictly positive number- strictly positive function $\tilde{s} : C^* \times C^* \to \mathbb{R}_+$, such that for all $(D, E) \in A$ the following representation holds:

$$(P \circ \tilde{C})(D, E) \quad = \quad \frac{\sum_{l=1}^{L+1} \tilde{s}(D, \tilde{C}_l^E) P^{\tilde{C}_l^E}}{\sum_{l=1}^{L+1} \tilde{s}(D, \tilde{C}_l^E)}$$

Moreover, $\tilde{s}(D \circ c, \tilde{C}_l^E) > \tilde{s}(D, \tilde{C}_l^E)$ for all $c \in \tilde{C}_l^E$.

Rough sketch of the proof:

The necessity part is straightforward calculation. The sufficiency part follows the rough structure of the proof of BGSS and Bleile (2014b), but differs in the crucial arguments. Again, the first step is to reason, why it is viable to transform $(P \circ \tilde{C})(D, E)$ for $(D, E) \in A$ into $(P \circ \tilde{f})(f_D, f_E)$ for appropriate adjusted definitions of $(f_D, f_E) \in A$.
The essential part of the proof is to derive the similarity weights $(s_i)_{i \leq m}$. This will be shown inductively over $div(f_D) \leq k \leq m$.

Step 1: Base case for the induction, take any triplet $\{i,j,k\} \subset \{1,..,m\}$ such that $f_D \in conv(\{(q^v)_{v\in\{i,j,k\}}\})$, i.e. $div(f_D) \leq 3$, and $(f_D, f_E) \in A$.
(i) For $(f_D, f_E) \in A$ such that $2 \leq div(f_D) \leq 3$ and $3 \leq div(f_E) \leq 4$ the categorization procedure is vanished and coincide with the BGSS framework.
(ii) For $(f_D, f_E) \in A$ such that $div(f_D) \leq 3$ and $div(f_E) \geq 4$.
(a) For $f_D \in (q^i, q^j)$ and $(f_D, f_E) \in A$, there exist some q^k such that admissibility condition (ii) holds. Then take the simplex spanned by $\{q^i, q^j, q^k\}$ and adopt Step 1 of BGSS, Bleile (2014a,b), i.e. find s_i, s_j, s_k, define $(P \circ \tilde{f})(f, f_E)$ and run the recursive procedure to cover all simplicial points on this simplex for the fixed f_E. This yields that for $(P \circ \tilde{f})_s(f_D, f_E) = (P \circ \tilde{f})(f_D, f_E)$ for all $f_D \in conv(\{q^i, q^j, q^k\})$ and fixed f_E.
(b) Since for an admissible pair $(f_D, f_E) \in A$ from (a) there might exist many k such that the admissibility condition (ii) holds. Repeat (a) for all such k.
(c) Since for $f_D \in (q^i, q^j)$ in (a) there exist many f_E such that (f_D, f_E) is admissible, repeat (a) and (b) for all these f_E.
(d) Repeat (a), (b) and (c) for any pair (q^i, q^j) with distinct $i, j \leq m$.
Thus, we have that $(P \circ \tilde{f})_s(f_D, f_E) = (P \circ \tilde{f})(f_D, f_E)$ for all admissible pairs (f_D, f_E) such that $div(f_D) \leq 3$.

Step 2: For all $(f_D, f_E) \in A$ such that $div(f_D) > 3$
By the properties of the categorization, we know that at least two f_E-categories contain a least two cases and another category contains at least one member, e.g. $\tilde{f}_1^{f_E} \supseteq \{q^1, q^i\}$, $\tilde{f}_2^{f_E} \supseteq \{q^2, q^j\}$, etc.
Let $f_D = \alpha q^j + (1-\alpha) f(j)$ (for some $\alpha \in (0,1)$) where $f(j)$ denotes the point in $conv\{(q^l)_{l\in\{1,\ldots,m\}\setminus j}\}$ that is on the line through f_D and q^j, as in BGSS.
Then the prototype based Concatenation Axiom and the induction assumption delivers

$$(P \circ \tilde{f})(f_D, f_E),\ (P \circ \tilde{f})_s(f_D, f_E) \in \bigcap_{t=1,2} ((P \circ \tilde{f})(q^t, f_E),\ (P \circ \tilde{f})(f(t), f_E)).$$

Applying the Collinearity Axiom yields the uniqueness of the intersection and the desired result.

Interpretation of Theorem

The theorem is described for general pairs of admissible databases $(D, E) \in A$. However, the most interesting and natural situation is given by identical information (D, D), which motivated our examination of a belief formation based on categorized information. For $(D, D) \in A$, an agent only employs the categorized information in database D and is not involved in the process of case based activation of categories as necessary for pairs $(D, E) \in A$. For (D, D), the admissibility condition for a "real" categorization of the information D requires that the database D is categorized at least into three different categories (which is in principle only possible for databases with at least five different cases). Thus, the most interesting and meaningful situation is for large databases which allow a sufficiently rich categorization structure.

A prototype based belief formation does not focus on employing the information contained in the most appropriate evoked (target) categories for a problem, but it takes into account the entire categorized information in a database. The belief is not only based on one target category, but across all categories. This tries to compensate for potential misassignments if the actual problem does not allow for a straightforward most appropriate category. However the process is not based on a detailed piece by piece evaluation of all cases and their induced estimates separately, but relies on the summarized coarse information on the category level. This is in line with the procedure in Anderson (1991). In particular, an agent only needs to compare and balance the categories (as an entity) at large and use their category specific predictions. Thus not all single pieces of information need to be evaluated, which is a severe (cognitive) simplification and captures the underlying aspect and motivation of a categorization (heuristic). The category based estimation $P^{\tilde{C}_l^D}$ is the main ingredient and eponymous for our belief, since it can be interpreted as representative or prototypical estimate associated with the category $\tilde{C}_l^D$. Each category has a unique representative prototypical estimate, which does not distinguish between between cases in the same category, i.e. for all $c \in \tilde{C}_l^E$ $(P \circ \tilde{C})(c, E) = P^{\tilde{C}_l^E}$. Implicitly, this means that a category is understood in terms of a prototypical element in the category that captures, compresses, aggregates, summaries and abstracts the essence and central tendency of a category (for prototype theory see e.g. Posner and Keele (1968), Reed (1972)). A specific representation of such an aggregated prototype based estimate is not implied, but a very natural prototype is simply the mean across previously experienced objects in the category. But also other statistics can serve as prototypes, such as min if the decision maker cares about worst case scenarios.

The weights $(\tilde{s}(\tilde{C}_l^D, D))$ that are assigned to each category specific estimate $P^{\tilde{C}_l^D}$ reflect the relevance of category l in the database D for the current problem. The weights do not only measure the relevance of the categories $\tilde{C}_l^D$, but also incorporate how often (in similarity weighted terms) this category is activated by the specific database D. [30] Thus, the specific content and structure (frequencies) of the (activating) databases are taken care of, e.g. two databases that are categorized in identical categories can induced different beliefs if they contain differently many cases of specific types, since then the relative relevance of the same categories can be altered.

In sum, a prototype based belief formation facilitates fast and cognitive less demanding predictions compared to "smoother" forms of similarity based reasoning as for instance provided by kernel-based predictions or BGSS, EG and Bleile (2014a). these approaches need to incorporate all single pieces of information. For the prototype based approach, an agent simply evaluates a problem in terms of prototypical thinking and reasoning, by averaging the categories' prototypical estimates, which can be derived, stored and retrieved solely on the category level, independent of the problem. A kernel based or BGSS belief requires a higher and more complex cognitive load and task by the need to store a large amount of information and generate many more (conditional) estimations that complicates the belief formation.

[30] Remember that a category does not contain repetitions of cases, i.e. $\tilde{C}_l^D = \tilde{C}_l^{D \circ c}$ for any $c \in \tilde{C}_l^D$.

7 Conclusion

This paper examines how beliefs are formed by agents that use a categorization procedure in order to process, store and retrieve the available information. The cognitive science literature emphasizes the role of categorical processing, thinking and reasoning. Based on this insight we axiomatize a two stage belief formation procedure in which agents employ categorized information and do not incorporate the available information piece by piece. We assume that an agent is equipped (or has acquired (evolutionary)) subconsciously with some intuition or heuristic how to (optimally) categorize the entire set of possible pieces of information. Based on such a naturally given categorization heuristics, we introduce a procedure that consistently categorizes databases. We consider two well known and observed procedures of categorizations, depending on how categories are activated for a new problem and how they are represented. One procedure relies only on the information in specific "target" categories that are the most appropriate categories for the current problem. Another procedure relies on all categories of the database, but that are not represented by their contained single pieces of information, but rest on so called prototypical elements that represent a summary or central tendency of the category. For both procedures, we require a minimum amount of complexity/diversity of the underlying information, such that an agent really engages into categorical processing and thinking. Otherwise an agent sticks just to piece by piece evaluation of the information.
The axioms on the belief level are closely related to the axioms introduced in BGSS and Bleile (2014b) and modified in a way to capture the categorization of information and their consequences for induced beliefs. The two versions of belief formation based on categorizations - category based and prototype based- are weighted sums of estimates induced by past categorized information. Whereas the category based belief relies on estimates of past observations in the target categories that are activated by the current problem, the prototype based belief relies on all category related prototypical estimates in the database. The weights that are assigned to the different estimates cover the similarity of the current problem with the single piece of information that induced the estimate or respectively its relevance with the particular category (or its prototype).
Compared to the beliefs axiomatized in BGSS, EG and Bleile (2014a), both belief formations based on categorized information reduce the cognitive effort extensively and thus are more realistic procedures for belief formations. For the category based belief an agent only needs to consider, evaluate and estimate each single piece of information within the target categories, i.e. only some subset of the available information. In the prototype based belief, an agent even thinks entirely categorical or in prototypes and thus treats information always on an aggregate level.

A Preparations for the Proofs

An essential step in the proofs will be to identify database with their frequency vectors, which allows to exploit the more tractable structure of the space of frequencies on C and to adopt the approach taken in BGSS (and use the mechanism of Bleile (2014a)). However, the proofs for categorization based beliefs formation requires some additional features, since in addition a categorization step is involved, which alters the crucial steps in the inductive proof

General Definitions for a Frequency Framework

We need to introduce some definitions regarding the frequency framework.

The set of all frequency vectors on the ordered set of basic cases $C = \{c^1, .., c^m\}$ is given by (since C is fixed we skip it in the following)
$\Delta(C) = \Delta := \{f = (f_1, ..., f_m) \text{ s. th. } f_i \in \mathbb{Q} \cap [0,1] \text{ for all } i \leq m \text{ and } \sum_{i \leq m} f_i = 1\}$
The following set represents all frequency vectors related to databases $D \in C^T$:

$$\Delta_T := \{f \in \Delta \; f_i = \frac{l_i}{T},\; l_i \in \mathbb{N}_+,\; \sum_{i=1}^{m} l_i = T \text{ and } \exists\, D \in C^T \text{ such that } f_D(c_i) = f_i = l_i/T\}$$

Observe that if $f \in \Delta_T(C)$, then $f \in \Delta_{TZ}(C)$ for all $Z \in \mathbb{N}_+$, i.e. the frequency vector f_D represents all databases D^Z for some $Z \in \mathbb{N}$ and we cannot relate it to any specific database D^Z for a specific $Z \in \mathbb{N}$.

Definition A.1
(i) O^m denotes the null-vector on $\mathbb{R}^m$.
(ii) For all $j \in \{1, 2, ..., m\}$ denote by q^j the j-th unit vector in $\mathbb{R}^m$, i.e. the frequency vector representing a database containing only case $c_j \in C$, i.e. $q^j = (0, ..., 0, \underbrace{1}_{j-th}, 0, ..., 0)^t$
(iii) For all $d \in \Delta$ its diversity is given by $div(d) := |\{i \leq m \,|\, d_i > 0\}|$

Definition A.2
(i) The $\subseteq$-relation on frequencies $\Delta \times \Delta$ is defined as follows for $d, e \in \Delta$:

$$d \subseteq e \textit{ if and only if } d_i \geq 0 \textit{ only if } e_i > 0 \textit{ for all } i \leq m.$$

(ii) Let $d \in \Delta_T$ and $e \in \Delta_L$, then the $\cap$-relation on $\Delta \times \Delta$ is defined by

$$d \cap e := \left(\left(\frac{\min\{d_i T, e_i L\}}{\sum_{i \leq m} \min\{d_i T, e_i L\}}\right)_{i \leq m}\right)$$ [31]

For definition (i) we have in mind that there exist T and L such that d represents a database of length T, i.e. $D \in C^T$ and e an $E \in C^L$ such that $\min_{j \leq m} d_j T \leq \min_{j \leq m} e_j L$.

B Proof of Theorem 5.1 (Category based belief formation)

It is straightforward to show that the representation satisfy the axioms.
The difficult part is the sufficiency direction, i.e. axioms imply representation. As before, the essential step in the proof will be to identify database with their frequency vectors, which allows to exploit the more tractable structure of the space of frequencies on C and to adopt the approach taken in BGSS (and use the mechanism of Bleile (2014a)). However, since in addition a categorization step is involved, the crucial steps in the inductive proof require different arguments.

[31] **(Consistency) Remark:** If $d \subseteq e$, then obviously $d \cap e = ((\frac{\min\{d_i T, e_i L\}}{\sum_{i \leq m} \min\{d_i T, e_i L\}})_{i \leq m}) = ((\frac{d_i T}{\sum_{i \leq m} d_i T})_{i \leq m}) = d$

B.1 Translating the database framework into frequencies

Why is it viable?

Remember, that we fixed a categorization structure $\tilde{C}$ and a problem $x \in X$.

In the following, we show that a consistent transformation form databases to frequencies is viable. Roughly, we want to identify a problem evoked categorization $\tilde{C}(D)$ of a database D by its frequency vector in $\Delta(C)$ such that $\tilde{C}(f_D) \in \Delta(C)$ corresponds to $\tilde{C}(D) \in C^*$ within the category based belief formation, i.e. such that $(P \circ \tilde{C})(D)$ corresponds to $(P \circ \tilde{C})(f_D)$. For this purpose, we exploit the structure of the categorization procedure and the axioms on the belief formation stage.
We need to show that the two stage procedure is independent of the order of the involved database and its length. However, as already discussed in Section 5.1, for the specific categorization procedure only the content matters and the order of cases is irrelevant, i.e. $\tilde{C}(D) = \tilde{C}(\pi(D))$ for any reordering $\pi(D)$ of the database D. Furthermore, the length of the database is immaterial, since the category based Concatenation Axiom implies that $(P \circ \tilde{C})(D^Z) = \sum_{i \le Z} \lambda_i (P \circ \tilde{C})(D) = (P \circ \tilde{C})(D)$ for all $Z \in \mathbb{N}$ and appropriate $\lambda \in \Delta^Z$. Consequently for a category based belief we can identify any database $D \in C^*$ by its frequency vector f_D, i.e. the category based belief translates from $(P \circ \tilde{C}) : C^* \to \Delta(R)$ to $(P \circ \tilde{C}) : \Delta(C) \to \Delta(R)$ by $(P \circ \tilde{C})(f_D) := (P \circ \tilde{C})(D)$.
We need to reformulate the categorization, axioms and results from databases to frequency vectors, given a fixed problem $x \in X$.

Categorization in frequency terms

Definition B.1
(i) Given a natural categorization $\tilde{C} = (\tilde{C}_1, .., \tilde{C}_L)$, the list of x-evoked categories on C, i.e. $\tilde{C}_x = (\tilde{C}_l)_{l \in M_x} \subseteq \tilde{C} \subseteq C^*$ for a corresponding $M_x \subseteq \{1, .., L\}$ translates to a x-evoked categorized frequency vector $\tilde{C}^x \in \Delta(C)$:

$$\tilde{C}^x \quad = \quad \Big(\frac{1}{|\tilde{C}_x|} \sum_{l \in M_x} 1_{\{\tilde{C}_l\}}(c^1), ..., \frac{1}{|\tilde{C}_x|} \sum_{l \in M_x} 1_{\{\tilde{C}_l\}}(c^m) \Big) \in \Delta(C)$$

which describes how often the ordered cases in C appear in the list of evoked categories.

(ii) A x-evoked categorization function for a database $D \in C^*$, i.e. $\tilde{C}(x, D) \in C^*$ translates to a x-evoked categorization function for a frequency vector $f = f_D$, i.e. $\tilde{C} : X \times \Delta(C) \to \Delta(C) \cup 0^m$ such that $\tilde{C}(x, f) = (\tilde{C}_1(x, f), ..., \tilde{C}_m(x, f)) \in \Delta(C) \cup 0^m$ is defined for $j \le m$ by

$$\tilde{C}_j(x, f) = f_j 1_{\{div(f) \le 3\}}(f) + \frac{f_j 1_{\{\tilde{C}^x\}}(q^j)}{\sum_{i \le m} f_i 1_{\{\tilde{C}^x\}}(q^i)} 1_{\{div(f) > 3\}}(f)$$

$\tilde{C}(x, f)$ is the frequency vector of a categorized database D (represented by frequency vector f) evoked by a problem x, i.e. describes the resulting frequencies of the cases in D that are also contained in the x-evoked categories.

Note, that $\tilde{C}(f_D)$ represents $\tilde{C}(E)$ for all $E = \pi(D^Z)$ and any $Z \in \mathbb{N}$.
As before, we will suppress a fixed x, i.e. $\tilde{C}(x, f) = \tilde{C}(f)$.

Axioms in frequency terms

Category based Concatenation Axiom
For all $T_i \in \mathbb{N}$ $(i = 1, 2)$ and any $f_i \in \Delta_{T_i}$, there exists $\lambda \in [0, 1]$, such that for $f = \frac{T_1}{T_1+T_2} f_1 + \frac{T_2}{T_1+T_2} f_2$

$$(P \circ \tilde{C})(f) = \lambda (P \circ \tilde{C})(f_1) + (1 - \lambda)(P \circ \tilde{C})(f_2),$$

where $\lambda = 0$ if and only if $\tilde{C}(f_1) = 0^m$.

Collinearity Axiom
No three of $\{((P \circ \tilde{C})(q^j))_{j \leq m}\}$ such that $\tilde{C}(q^j) \neq 0^m$ are collinear.

B.2 Theorem 5.1, sufficiency part in frequency terms

Proposition B.1
Let there be a function $(P \circ \tilde{C}) : \Delta(C) \to \Delta(R)$, where $P : \Delta(C) \to \Delta(R)$ and $\tilde{C} : \Delta(C) \to \Delta(C)$ a categorization function on the set of frequency vectors.
If $(P \circ \tilde{C})$ satisfies the category based Concatenation and Collinearity Axiom, then there exist unique probability vectors $(P^j)_{j \leq m} \in \Delta(R)$, and unique -up to multiplication by a strictly positive number- strictly positive numbers $(s_j)_{j \leq m} \in \mathbb{R}$, such that for all $q \in \Delta(C)$ such that $\tilde{C}(q) \neq 0^m$

$$(P \circ \tilde{C})(q) = \frac{\sum_{j \leq m} s_j \tilde{C}_j(q) P^j}{\sum_{j \leq m} s_j \tilde{C}_j(q)} \qquad (2)$$

where $\tilde{C}_j(q)$ denotes the frequency of case c_j in $\tilde{C}(q)$.

B.3 Proof of Theorem 5.1, sufficiency part in frequency terms

Step 0:
Obviously, by the definition of the categorization on frequency vectors we need to choose $P^j = (P \circ \tilde{C})(q^j)$ for $j \leq m$, since $\tilde{C}_j(q) \in \{q^j, 0^m\}$ and for some q (e.g. for q s.th. $div(q) \leq 3$)) $\tilde{C}_j(q) = q^j$.

The aim is to find numbers $(s_j)_{j \leq m}$ such that representation (2) holds for all $q \in \Delta(C)$.

Step 1: $|C| = m = 3$

By the definition of the categorization function we have that $\tilde{C}(q) = q$ for all $q \in \Delta(C)$, since $div(q) \leq 3$. In such a situation a categorization of information does not take place and thus the framework coincides with the one in BGSS. Basically $(P \circ \tilde{C})(q)$ coincides with $P(q)$ in the BGSS framework for $div(q) \leq 3$ and therefore Step 1 of the proof in BGSS can be directly adopted for these frequency vectors.

Step 2: Now we consider $|C| = m > 3$.
Step 2.1: Defining the similarity weights

Using the considerations from Step 1 for all triplets $\{i, j, k\}$ and $C = \{c^i, c^j, c^k\}$ we can derive the similarity weights $(s_v^{\{i,j,k\}})_{v\in\{i,j,k\}}$ and we know that for all $q \in \Delta(\{q^i, q^j, q^k\})$, the following representation holds

$$(P^{\{i,j,k\}} \circ \tilde{C})(q) = \frac{\sum_{v\in\{i,j,k\}} s_v^{\{i,j,k\}} \tilde{C}_v(q) P^v}{\sum_{v\in\{i,j,k\}} s_v^{\{i,j,k\}} \tilde{C}_v(q)},$$

where for all $v \in \{i, j, k\}$ P^v are independent of the triplet $\{i, j, k\}$ (by Step 0) and $(s_v^{\{i,j,k\}})_{v\in\{i,j,k\}}$ are unique up to multiplication by a positive number.

With a similar reasoning as in the proof in BGSS Step 2.1 and using again that $\tilde{C}(q) = q$ for q such that $div(q) \leq 3$, we can show that the similarity values $s_v^{\{i,j,k\}}$ are independent of the choice of i, j and k for all v.
Thus, given these $(s_v)_{v\leq m}$ we can define for all $q \in \Delta$

$$(P \circ \tilde{C})_s(q) := \frac{\sum_{j\leq m} s_j \tilde{C}_j(q) P^j}{\sum_{j\leq m} s_j \tilde{C}_j(q)}.$$

Obviously, $(P \circ \tilde{C})_s$ satisfies the category based Concatenation Axiom.

Step 2.2: Completion to all $q \in \Delta(C)$, i.e. show that for all $q \in \Delta(C)$ $(P \circ \tilde{C})_s(q) = (P \circ \tilde{C})(q)$

We proof this by induction over k for $q \in \Delta(C)$ such that $div(q) = k$.
By Step 1 we know that the claim $(P \circ \tilde{C})_s(q) = (P \circ \tilde{C})(q)$ is true for all $q \in \Delta(C)$ such that $div(q) \leq 3$. This serves as the base case for the induction. Now we assume that $(P \circ \tilde{C})_s(q) = (P \circ \tilde{C})(q)$ for $q \in \Delta(C)$ such that $div(q) = k - 1$ and we will show it for $q \in \Delta(C)$ such that $div(q) = k$

A similar construction as in BGSS, but with different reasoning yields the result. Let $q = \sum_{l\in K} \alpha_l q^l$ with $\alpha_l > 0$ and $K \subseteq \{1, ..., m\}$ such that $|K| = k$.
Define for all $l \in K$ the frequency vector $q(l)$ to be the vector in $conv(\{(q^j)_{j\in K\setminus l}\})$ such that q lies on the line generated by $(q(l), q^l)$. By the category based Concatenation Axiom there exists some $\lambda \in [0, 1]$ such that for all $j \in K$

$$(P \circ \tilde{C})(q) = \lambda(P \circ \tilde{C})(q^j) + (1 - \lambda)(P \circ \tilde{C})(q(j)).$$

We distinguish between two cases:

(i) If there exists a $j \in K$ such that $q^j \notin \tilde{C}(q)$, then the category based Concatenation Axiom implies for $q = \alpha q^j + (1 - \alpha)q(j)$ for an appropriate $\alpha \in (0, 1)$ that $(P \circ \tilde{C})(q) =$

$(P \circ \tilde{C})(q(j))$, since $(P \circ \tilde{C})(q^j)$ receives zero-weight. The same is true for $(P \circ \tilde{C})_s$, since it satisfies also the category based Concatenation Axiom. However, since $div(q(j)) = k - 1$, the induction assumption yields $(P \circ \tilde{C})_s(q(j)) = (P \circ \tilde{C})(q(j))$ and we get the desired result:

$$(P \circ \tilde{C})(q) = (P \circ \tilde{C})(q(j)) = (P \circ \tilde{C})_s(q(j)) = (P \circ \tilde{C})_s(q).$$

(ii) If for all $j \in K$ $q^j \in \tilde{C}(q)$, then there are $k > 3$ many q^j such that $\tilde{C}(q^j) \neq 0^m$. Again, the category based Concatenation Axiom applied to the k many decompositions of $q = \alpha^j q^j + (1 - \alpha^j)q(j)$ for appropriate $\alpha^j \in (0, 1)$ for all $j \in K$ yields that $(P \circ \tilde{C})(q)$ lies in the interior of the intervals spanned by $((P \circ \tilde{C})(q^j), (P \circ \tilde{C})(q(j)))$ for all $j \in K$. Since for no three different $j \in K$ these intervals can lie on the same line by the Collinearity Axiom (since no three of $\{((P \circ \tilde{C})(q^j) = P^j)_{j \leq C}\}$ are collinear), there must exist some intersections of the lines. However, since $(P \circ \tilde{C})(q)$ lies in all these intervals, the intersection must be unique and exactly equal to $(P \circ \tilde{C})(q)$. However, also $(P \circ \tilde{C})_s(q)$ lies on all these intervals, since by induction assumption $(P \circ \tilde{C})(f) = (P \circ \tilde{C})_s(f)$ for all $f \in \Delta(C)$ such that $div(f) \leq k - 1$ (which q^j and $q(j)$ satisfy). Thus the unique intersection can only be $(P \circ \tilde{C})_s(q)$, which shows the equivalence of $(P \circ \tilde{C})(q) = (P \circ \tilde{C})_s(q)$. This completes the proof for $|C| > 3$ and eventually the Proposition B.1 and Theorem 5.1. □

C Proof of Theorem 6.1 (Prototype based belief formation)

It is straightforward to show that the representation satisfies the axioms.
To show that the axioms imply the representation requires some work. As before, we identify database with their frequency vectors in order to adopt the approach taken in BGSS (and use the mechanism of Bleile (2014a)). However, the additional categorization procedure alters the reasoning in the inductive proof significantly.

C.1 Translating the database framework into frequencies

Why is it viable?

We want to will identify the prototype based belief induced by categorized databases, i.e. $(P \circ \tilde{C}) : C^* \times C^* \to \Delta(R)$ by a belief $(P \circ \tilde{f}) : \Delta(C) \times \Delta(C) \to \Delta(R)$ based on frequency vectors and their induced categorization structures $\tilde{f} \in P(\Delta(C))$, i.e. $\tilde{C}$ is represented by $\tilde{f}$ and $(P \circ \tilde{C})(D, E)$ by $(P \circ \tilde{f})(f_D, f_E)$.
Let $(D, E) \in A$.
(1) In a first step we exploit Remark 6.2, i.e. that $(P \circ \tilde{C})(D, E) = (P \circ \tilde{C})(\pi(D), \pi'(E))$ where π, π' are permutations that reorder the cases in D and respectively in E arbitrarily. Basically it says that orders of databases are totally immaterial for the induced prototype based belief, i.e. only frequency vectors matter.
(2) a) The definition of a prototype based belief and database related categorization structures yields

$$(P \circ \tilde{C})(D, E) = P(\tilde{C}^E(D)) = P(\tilde{C}^{E^Z}(D)) = (P \circ \tilde{C})(D, E^Z) \text{ for all } Z \in \mathbb{N}$$

b) In addition the prototype based Concatenation Axiom implies for $D^Z = D \circ ... \circ D$ for all $Z \in \mathbb{N}$

$$(P \circ \tilde{C})(D^Z, E) = (P \circ \tilde{C})(D, E)$$

Combining (2)a) and (2)b) yields for all $(D, E) \in A$ and $V, Z \in \mathbb{N}$

$$(P \circ \tilde{C})(D^V, E^Z) = (P \circ \tilde{C})(D, E) \tag{3}$$

and thus also the lengths of the involved databases are immaterial as well.
Combining (1) and equation (3) shows that we can identify any $D, E \in C^*$ by their frequency vectors $f_D, f_E \in \Delta(C)$ for a prototype based belief formation.

Consequently, we can rewrite our framework into a frequency framework.

Categorization structures in frequency terms

Definition C.1
A categorization structure $\tilde{f} = \{\tilde{f}_1, ..., \tilde{f}_L\}$ on $\Delta(C)$ satisfies the following properties
(i) $(\tilde{f}_l)_i := \frac{1}{|\tilde{C}_l|} 1_{\tilde{C}_l}(c_i)$ for all $l \leq L$ and $i \leq m$
(ii) $(\tilde{f}_k)_i > 0$ if and only if $(\tilde{f}_j)_i = 0$ for all $i \leq m$ and any distinct $j, k \leq L$ (i.e. disjointness)
(iii) $\sum_{l \leq L} (\tilde{f}_l)_i > 0$ for all $i \leq m$ (i.e. all cases are categorized)
(iv) for all $l \leq L$ $(\tilde{f}_l)_i \leq 1/2$ and $\tilde{f}_L(i) \leq 1$ for all $i \leq m$ (i.e. specific content structure)

Define for all $l \leq L+1$ and any $q \in \Delta(C)$ $A(l, q) := \{j \leq m | (\tilde{f}_l)_j > 0 \text{ and } q_j > 0\}$, i.e. the indices j such that case q^j contained in database q belongs to category l $\tilde{f}_l$.

Definition C.2
Let $\tilde{f} = \{\tilde{f}_1, ..., \tilde{f}_L\}$ be a categorization structure on $\Delta(C)$. A categorization structure $\tilde{f}^q = \{\tilde{f}_1^q, ..., \tilde{f}_{L+1}^q\}$ on q is defined by
(i) if $div(q) \leq 4$, i.e. $q \in conv(\{q^h, q, q^j, q^k\})$, then $\tilde{f}^q = \{\tilde{f}_1 = q^h, \tilde{f}_2 = q^i, \tilde{f}_3 = q^j, \tilde{f}_4 = q^k, \tilde{f}_5 = 0^m, ..., \tilde{f}_{L+1} = 0^m\}$

(ii) if $div(q) > 4$, then $\tilde{f}^q$ is given for all $l \leq L+1$ by

$$\begin{aligned} \tilde{f}_l^q : &= \begin{cases} \left(\frac{1_{A(l,q)}(1)}{|A(l,q)|}, ..., \frac{1_{A(l,q)}(m)}{|A(l,q)|}\right) & \text{if } |A(l,q)| \geq 2 \\ \emptyset = 0^m & \text{if } |A(l,q)| \leq 1 \end{cases} \\ \tilde{f}_{L+1}^q &= \frac{\sum_{l \leq L} 1_{\{l \leq L || A(l,q)| = 1\}}(l) q^{A(l,q)}}{\sum_{l \leq L} 1_{\{l \leq L || A(l,q)| = 1\}}(l)} \end{aligned}$$[32]

By the equation (1) it is also possible to define $\tilde{C}^E(D)$ in frequency term.

Definition C.3
For any $q, e \in \Delta(C)$ such that $q_i \geq 0$ only if $e_i > 0$:
$\tilde{f}^e(q) = \{\tilde{f}_1^e 1_{\{q | \exists i \leq m \; q_i > 0 \text{ and } \tilde{f}_1^e(i) > 0\}}(q), ..., \tilde{f}_{L+1}^e 1_{\{q | \exists i \leq m \; q_i > 0 \text{ and } \tilde{f}_{L+1}^e(i) > 0\}}(q)\} \in P(\Delta(C))$
is the result of an q-induced e-categorization function $\tilde{f} : \Delta(C) \times \Delta(C) \to P(\Delta(C))$.

Admissibility condition in frequency terms

Definition C.4

Define the set of admissible pairs $(q,e) \in \Delta(C) \times \Delta(C)$ *by the following*

$$A := \Big\{(q,e)| \quad (i)\ \textit{for all}\ i \le m\ q_i > 0,\ \textit{only if}\ e_i > 0$$
$$(ii)\ \textit{if}\ div(q) = 2\ \textit{then}\ \exists i \le m\ q_i = 0\ \textit{and}\ e_i > 0$$
$$\textit{s. th. for}\ \alpha \in (0,1)\ |\tilde{f}^e(\alpha q + (1-\alpha)q^i)| = 3$$
$$(iii)\ \textit{if}\ div(q) \ge 3\ \textit{then}\ |\tilde{f}^e(q)| \ge 3\Big\}$$

Axioms in frequency terms

Prototype based Concatenation Axiom

For all $(\alpha q + (1-\alpha)q', e) \in A$ for some $\alpha \in (0,1)$, there exists $\lambda \in (0,1)$

$$(P \circ \tilde{f})(\alpha q + (1-\alpha)q', e) = \lambda (P \circ \tilde{f})(q,e) + (1-\lambda)(P \circ \tilde{f})(q', e)$$

Identity Axiom

For all $(q,e), (q',e) \in A$ such that there exists a unique e-category $\tilde{f}^e_l(q) \neq 0^m$ and $\tilde{f}^e_l(q') \neq 0^m$ and $\tilde{f}^e_j(q) = 0^m = \tilde{f}^e_j(q)$ for all $j \neq l \le L+1$ it holds

$$(P \circ \tilde{f})(q_1, e) = (P \circ \tilde{f})(q_2, e)$$

Collinearity Axiom

No three distinct vectors of $\{((P \circ \tilde{f})(q^j, e))_{j \le m}\}$ are collinear for any $(q^j, e) \in A$

C.2 Theorem 6.1, sufficiency part in frequency terms

Proposition C.1

Let there be given a function $(P \circ \tilde{f}) : \Delta(C) \times \Delta(C) \to \Delta(R)$*, where* $\tilde{f}$ *a categorization function* $\tilde{f} : \Delta(C) \times \Delta(C) \to P(\Delta(C))$ *with related categorization structure* $\tilde{f}^e$ *for all* $e \in \Delta(C)$*. If the prototype based belief* $(P \circ \tilde{f})$ *satisfies the prototype based Concatenation, Identity and the Collinearity Axiom, then*
for all $e \in \Delta(C)$ *and each* $l \le L+1$ *such that* $\tilde{f}^e_l \neq 0^m$ *there exist* $P^{\tilde{f}^e_l} \in \Delta(R)$ *and a strictly positive -and unique up to multiplication with a strictly positive number- values* $s = (s_j)_{j \le m}$ *such that for all admissible* $(q,e) \in A$

$$(P \circ \tilde{f})(q,e) = \frac{\sum_{l \le L+1} \tilde{s}(\tilde{f}^e_l, q) P^{\tilde{f}^e_l}}{\sum_{l \le L+1} \tilde{s}(\tilde{f}^e_l, q)} \qquad (4)$$

where $\tilde{s}(\tilde{f}^e_l, q) := \sum_{\{i \le m | q^i \subseteq q \text{ and } q^i \subseteq \tilde{f}^e_l\}} q_i s_i$

C.3 Proof of Theorem 6.1, sufficiency part in frequency terms

Step 0:

We get directly that for all $l \le L+1$ and all $e \subseteq \Delta(C)$ $P^{\tilde{f}^e_l}$ must be chosen by $(P \circ \tilde{f})(q^j, e)$ for an appropriate $j \le m$ such that $q^j \subseteq \tilde{f}^e_l$. By the Identity Axiom this is unique.

The aim is to find the $(s_j)_{j\leq m}$ such that the representation (4) holds for all admissible pairs $(q,e)\in A$. We proceed in two steps, where the first step considers $(q,e)\in A$ such that $div(q)\leq 3$ and in a second step we inductively generalize it to q with larger diversity.

Step 1: $(q,e)\in A$ **such that** $div(q)\leq 3$**,** i.e. take any triplet $\{i,j,k\}\subset\{1,..,m\}$ such that $q\in conv(\{(q^v)_{v\in\{i,j,k\}}\})$.

Step 1.1: $(q,e)\in A$ **such that** $div(q)=1$ is covered in Step 0.

Step 1.2: $(q,e)\in A$ **such that** $2\leq div(q)\leq 3$ **and** $3\leq div(e)\leq 4$

Note that (q,e) with $div(e)=2$ are not admissible.
By the definition of an e-categorization structure the categorization step vanishes for all such pairs (q,e). This means that the prototype based framework (and axioms) directly amounts to the BGSS framework, i.e. $(P\circ\tilde{f})(q,e)$ coincides with the belief $P(q)$ in BGSS. Applying their proof yields the desired representation for all (q,e) with the above properties.

Step 1.3: $(q,e)\in A$ **such that** $div(q)\leq 3$ **and** $div(e)>4$

Step 1.3.1: Determine the similarity weights $(s_v)_{v\in\{i,j,k\}}$

Define for all triplets $\{i,j,k\}\subset\{1,..,m\}$ $q^*_{\{i,j,k\}}:=\frac{1}{3}(q^i+q^j+q^k)$
Obvious $div(q^*_{\{i,j,k\}})=3$ and to fulfill the admissability take e such that $|\tilde{f}^e(q^*_{\{j,k,l\}})|=3$ and hence each q^v for $v\in\{i,j,k\}$ needs to be contained in a different category of $\tilde{f}^e$.
We **assume for convenience that** $q^v\in\tilde{f}^e_v$ (it simplifies the notational effort extensively) for $v\in\{i,j,k\}$ which is possible after renaming the categories appropriately.
Observe that, since $(q^*_{\{i,j,k\}},e)$ is admissible, also $(q,e)\in A$ for any $q\subseteq e$ and $q\in conv(\{q^i,q^j,q^k\})$.

Now, we use $q^*_{\{i,j,k\}}$ to determine the values $(s_v)_{v\in\{i,j,k\}}$ given in the representation of the theorem. By the prototype based Concatenation Axiom there exist $\lambda=(\lambda_v)_{v\in\{i,j,k\}}\in int(\Delta^3)$ such that

$$(P\circ\tilde{f})(q^*_{\{i,j,k\}},e)=\sum_{v\in\{i,j,k\}}\lambda_v P^{\tilde{f}^e_v},$$

where we have used that $(P\circ\tilde{f})(q^v,e)=P^{\tilde{f}^e_v}$, as shown in Step 0.
The representation of the theorem (plugging in the definition of $\tilde{s}$) delivers

$$(P\circ\tilde{f})(q^*_{\{i,j,k\}},e)=\frac{\sum_{v\in\{i,j,k\}}s_v 1/3 P^{\tilde{f}^e_v}}{\sum_{v\in\{i,j,k\}}s_v 1/3}$$

Equating these two equations and using that $(P^{\tilde{f}^e_v})_{v\in\{i,j,k\}}$ are not collinear yields a solution for $(s_v)_{v=i,j,k}$, (i.e. $s_v = \frac{\lambda_v}{\sum_{v=i,j,k}\lambda_v}$). These s_v might depend on the specific triplet used in $q^*_{\{i,j,k\}}$, whereas obviously $P^{\tilde{f}^e_v} = (P\circ\tilde{f})(q^v, e)$ is independent of $\{i, j, k\}$, since only depending on e and q^v. However, similar as in Step 2.1 of the previous proof (or as in BGSS) we can show that the similarity values can be chosen independently of the particular triplet.

Thus, given these $s = (s1, ..., s_m)$ we define for any triplet $\{i, j, k\} \subset \{1, ..., m\}$ and $q \in conv(\{q^i, q^j, q^k\})$ and e such that $(q, e) \in A$ the following prototype based belief

$$(P\circ\tilde{f})_s(q, e) := \frac{\sum_{v\in\{i,j,k\}} s_v q_v P^{\tilde{f}^e_v}}{\sum_{v\in\{i,j,k\}} s_v q_v} \tag{5}$$

Recall that we assumed for convenience that $q^v \in \tilde{f}^e_v$, which allows this easy representation. Furthermore, observe that P_s satisfies the prototype based Concatenation Axiom.

Step 1.3.2: Completion to all $(q, e) \in A$ such that $div(q) \leq 3$

Define $E := \{(q, e) \in A \mid (P\circ\tilde{f})_s(q, e) = (P\circ\tilde{f})(q, e)\}$. In the following we want to derive that for all $q \in conv(\{q^i, q^j, q^k\})$ and all e such that $(q, e) \in A$ also $(q, e) \in E$.

The idea is to partition the simplex $\Delta(\{q^i, q^j, q^k\})$ recursively into sub-triangles-so called simplicial partitions. A recursive simplicial partition is based on halving the lines between the vertices of the triangles resulting form a previous partitioning. The vertices of the particular triangles are called the simplicial points (of a t-th simplicial partition). Starting from the simplex, the first simplicial partition results in 4 sub-triangles by connecting the middle points between two of $\{q^i, q^j, q^k\}$. The second simplicial partition consist of 16 smaller triangles resulting from applying the same "halving"-procedure to each of the 4 subtriangles in the first simplicial partition, and so forth.

As in BGSS, the idea is to show that all simplicial points g of any simplicial partition such that $(g, e) \in A$ it also holds that $(g, e) \in E$. We apply the mechanism as in Bleile (2014a,b) to cover all simplicial points recursively. Then it is possible to cover each $q \in cov(\{q^i, q^j, q^k\})$ by sequences of (appropriate) simplicial points.

In order to illustrate the intuition, we will describe only the first step, i.e. the simplicial points of the first simplicial partition. The further steps are analogously modified versions (to the prototype based setup) of arguments used in Bleile (2014b).

Remark C.1

For all $e \in \Delta(C)$ such that $q^i \subset e$ the pair (q^i, e) is admissible. However for q such that $div(q) \in \{2, 3\}$ this does not hold true in general. Nevertheless the sets of frequency vectors $e \in \Delta(C)$ that makes (q, e) admissible for $div(q) = 3$ coincides with the set of frequency vectors that make (q', e) admissible for $q' \subset q$ and $div(q') = 2$. This follows directly from the definition of the admissibility condition (ii). This guarantees there exists a common set of vectors e that make any t-th ($t \geq 1$) simplicial points admissible, when it is paired with such an e. This is important for our recursive procedure of combining all (differently diverse) simplicial points (as well be clear below).

Consider q such that $2 \leq div(q) \leq 3$

Step A:

Consider any two distinct q^i, q^j $(i, j \leq m)$, w.l.o.g let $i = 1, j = 2$. Define $q_1^1 := \frac{1}{2}q^1 + \frac{1}{2}q^2$ and obviously $div(q_1^1) = 2$. For all e such that $(q, e) \in A$, there exist some q^k (which might be different for different e) such that for $\alpha \in (0,1)$ $|\tilde{f}^e(\alpha q_1^1 + (1-\alpha)q^k)| = 3$. q_1^1 is a simplicial point of the first simplicial partition of the triangle spanned by $conv(\{q^1, q^2, q^k\})$. The prototype based Concatenation Axiom delivers the existence of $\beta, \gamma \in (0,1)$ such that

$$\begin{aligned}(P \circ \tilde{f})(q^*_{\{1,2,k\}}, e) &= \beta(P \circ \tilde{f})(q_1^1, e) + (1-\beta)(P \circ \tilde{f})(q^k, e) \\ (P \circ \tilde{f})(q_1^1, e) &= \gamma(P \circ \tilde{f})(q^1, e) + (1-\gamma)(P \circ \tilde{f})(q^2, e)\end{aligned}$$

Hence we get (using the notation that (a,b) indicates the line running through a and b)

$$(P \circ \tilde{f})(q_1^1, e) \in ((P \circ \tilde{f})(q^1, e), (P \circ \tilde{f})(q^2, e)) \cap ((P \circ \tilde{f})(q^k, e), (P \circ \tilde{f})(q^*_{\{1,2,k\}}, e)) \qquad (6)$$

The same holds true for $(P \circ \tilde{f})_s$, since it also satisfies the prototype based Concatenation Axiom. Moreover, by Step 0 we know already that for all $(q^j, e) \in A$, also $(q^j, e) \in E$ and that for any triplet $\{i, j, k\}$ and all $(q^*_{\{i,j,k\}}, e) \in A$ also $(q^*_{\{i,j,k\}}, e) \in E$.

Thus, it only remains to check, whether the induced prototype based beliefs of $(q^1, e), (q^2, e), (q^k, e)$ and $q^*_{\{1,2,k\}}$ are not collinear. Then, the two lines involved in (6) have an unique intersection. This implies then that $(P \circ \tilde{f})(q_1^1, e)$ and $(P \circ \tilde{f})_s(q_1^1, e)$ must coincide, since both lie on both lines, i.e. $(q_1^1, e) \in E$. However, the non-collinearity can be easily seen, since (q^v, e) for $v = 1, 2, k$ induce three different prototype based beliefs $P^{\tilde{f}_v^e}$ (since $|\tilde{f}^e(\alpha q_1^1 + (1-\alpha)q^k)| = 3$ for any $\alpha \in (0,1)$) that are not collinear by the Collinearity Axiom.

Analogously the other simplicial points of the first simplicial partition of $conv(\{q^1, q^2, q^k\})$ are analyzed, i.e. for $q_1^2 = \frac{1}{2}(q^1 + q^k)$ and $q_1^3 = \frac{1}{2}(q^2 + q^k)$. Following the (slightly modified) reasoning/method as in Bleile (2014a,b) Step 1, one can show that for the chosen e, all simplicial points of any t-th simplicial partition are in E. Finally, one can find a sequence of simplicial points that converges to any $q \in \Delta(\{q^i, q^j, q^k\})$, where also it induced prototype based belief converges. (for the details see the proofs of BGSS or the above mentioned sections in Bleile). Thus for a given e (which induces a q^k as above) $(q, e) \in E$ for all $q \in conv(\{q^1, q^2, q^k\})$ such that $(q, e) \in A$.

Step B:

Observe that for each admissible pair $(q, e) \in A$ such that $div(q) = 2$ the frequency vector e induces a set of frequency vectors $e_{ad} := \{q^k \mid \exists k \leq m \text{ s. th. } |\tilde{f}^e(\alpha q_1^1 + (1-\alpha)q^k| = 3\}$. Hence keeping q^1, q^2 fixed, the same procedure as in Step A can be applied to the triangle $conv(\{q^1, q^2, q^k\})$ for all $k \leq m$ such that $q^k \in e_{ad}$ for this specific e.

Step C:

Now, apply Step A and B to all e such that $(q, e) \in A$ for $div(q) \in \{2, 3\}$

Step D:

Finally, applying the procedure to all possible pairs q^i, q^j instead of $i = 1, j = 2$ we get that for all admissible $(q, e) \in A$ such that $div(q) \leq 3$ the claim $(q, e) \in E$, i.e. $A = E$ for all q such $div(q) \leq 3$, which concludes the proof of Step 1.

Now, we consider the situation for $(q, e) \in A$ such that $div(q) > 3$. Therefore we need an extended definition of $(P \circ \tilde{f})_s$ for $(q, e) \in \Delta(C) \times \Delta(C)$:

$$(P \circ \tilde{f})_s(q, e) := \frac{\sum_{i \leq m} s_i q_i (\sum_{l \leq L+1} 1_{\tilde{f}_l^e}(q^i) P^{\tilde{f}_l^e})}{\sum_{i \leq m} s_i q_i}$$

The indicator function appears in comparison to definition (5) since it is not clear to which category $\tilde{f}_l^e$ a specific unit vector q^i belongs. Basically $(P \circ \tilde{f})_s$ is a reformulation of the representation (4)

Step 2: Show that $(q, e) \in E$ for all $(q, e) \in A$ such that $div(q) > 3$

We prove inductively that for all $m \leq k \geq 3$ with $div(q) = k$ and all admissible pairs $(q, e) \in A$ it holds that $(P \circ \tilde{f})_s(q, e) = (P \circ \tilde{f})(q, e)$, i.e. $(q, e) \in E$.
We take the situation $div(q) = k = 3$, which was shown in Step 1, as the basis of the induction and assume that the claim is true for all $(q, e) \in A$ such that $div(q) = k - 1$ for $k > 3$.
Take any $(q, e) \in A$ such that $div(q) = k$, w.l.o.g. $q \in int(conv(\{q^1, .., q^k\}))$. By admissibility $|\tilde{f}^e(q)| \geq 3$ holds, i.e. there are at least two categories $\tilde{f}_l^e$ for some $l \leq L + 1$ containing at least two different cases q^j for some $j \leq m$ and another category containing at least one different case. W.l.o.g. let these categories be given in the following way (for distinct $i, j, k \neq 1, 2$)

$$\begin{aligned} \tilde{f}_1^e(q) &\supseteq \{q^1, q^i\} \\ \tilde{f}_2^e(q) &\supseteq \{q^2, q^j\} \\ \tilde{f}_{L+1}^e(q) &\supseteq \{q^k\} \\ \tilde{f}_l^e(q) &\supseteq \emptyset \text{ for all } l \neq 1, 2 \leq L \end{aligned}$$

Now, let q be decomposed by $q = \alpha_t q^t + (1 - \alpha_t) q(t)$, where for $t = 1, 2$ $q(t)$ is the point in $conv(\{q^j | j \leq k, j \neq t\})$ that is on the line connecting q^t and q and $\alpha_t \in (0, 1)$ accordingly. By the prototype based Concatenation Axiom we know that there exist $\lambda_t \in (0, 1)$ such that

$$(P \circ \tilde{f})(q, e) = \lambda_t (P \circ \tilde{f})(q^t, e) + (1 - \lambda_t)(P \circ \tilde{f})(q(t), e)$$

This is possible since $(q^t, e) \in A$ and $(q(t), e) \in A$, since $\tilde{f}^e(q(t)) = \tilde{f}^e(q)$ by construction. Since in addition $(q^t, e), (q(t), e) \in E$ by the induction assumption (since $div(q(t)) = k - 1$) and $(P \circ \tilde{f})_s$ satisfies the prototype based Concatenation Axiom, we have

$$(P \circ \tilde{f})(q, e),\ (P \circ \tilde{f})_s(q, e) \in \bigcap_{t=1,2} ((P \circ \tilde{f})(q^t, e),\ (P \circ \tilde{f})(q(t), e))$$

This intersection is unique since for $t = 1, 2$ it can be shown that the following holds:

$$(P \circ \tilde{f})(q^2, e) = P^{\tilde{f}_2^e} \notin ((P^{\tilde{f}_1^e}, (P \circ \tilde{f})(q(t), e)) =: h \tag{7}$$

If this would not be true, i.e. if $P^{\tilde{f}_2^e}$ would be on this line h, it would require that $(P \circ \tilde{f})(q(t), e)$ is on the line between $P^{\tilde{f}_1^e}$ and $P^{\tilde{f}_2^e}$. However by construction $(P \circ \tilde{f})(q(t), e) \in int(conv(\{P^{\tilde{f}_1^e}, P^{\tilde{f}_2^e}, P^{\tilde{f}_k^e}, ...\}))$, which implies that it cannot lie on $(P^{\tilde{f}_1^e}, P^{\tilde{f}_2^e})$, since by the Collinearity Axiom no three of $(P^{\tilde{f}_l^e})_l$ are collinear. Thus in sum, claim (7) is true, which implies that for $t = 1, 2$ the lines based on $((P \circ \tilde{f})(q^t, e), (P \circ \tilde{f})(q(t), e))$ are distinct and intersect uniquely in $(P \circ \tilde{f})(q, e) = (P \circ \tilde{f})_s(q, e)$, i.e. $(q, e) \in E$.
This completes the entire proof. □.

References

[1] Al-Najjar, N. I. and Pai, M. (2014). Coarse decision making and overfitting. *Journal of Economic Theory, 150, 467-486.*

[2] Allport, G.W. (1954). The Nature of Prejudice. *Addison Wesley, Cambridge.*

[3] Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review, 98-3, 409-429.*

[4] Bargh, J.A. (1994). The Four Horsemen of Automaticity: Awareness, Intention, Efficiency, and Control in Social Cognition. *in Handbook of Social Cognition, 1, 1-40.*

[5] Bargh, J.A. (1997). The Automaticity of Everyday Life. *in The Automaticity of Everyday Life: Advances in Social Cognition, 10, 1-61.*

[6] Bargh, J.A. (1999). The Cognitive Monster: The Case Against the Controllability of Automatic Stereotype Effects. *in Dual Process Theories in Social Psychology, 361-382.*

[7] Billot, A. Gilboa, I., Samet, D., Schmeidler, D. (2005). Probabilities as Similarity-Weighted Frequencies. *Econometrica, 73, 1125-1136.*

[8] Bleile, J. (2014a). Cautious Belief Formation. *Working Paper No.507, Center for Mathematical Eonomics, Bielefeld.*

[9] Bleile, J. (2014b). Limited attentioon in case based beleif formation. *Working Paper, Center for Mathematical Eonomics, Bielefeld.*

[10] Eichberger J., Guerdjikova, A. (2010). Case-Based Belief Formation under Ambiguity. *Mathematical and Social Sciences, 60-3, 161-177.*

[11] Fryer, R., Jackson, M. O. (2008). A categorical model of cognition and biased decision making. *The B.E. Journal of Theoretical Economics (Contributions), 8-1, 1-42.*

[12] Gärdenfors, P. (2000). Conceptual Spaces: The Geometry of Thought. *MIT Press, Cambridge.*

[13] Gelman, S. A., Markman, E. M. (1986). Categories and induction in young children. *Cognition, 23, 183-209.*

[14] Gilboa I., Schmeidler, D. (1995). Case-Based Decision Theory *Quarterly Journal of Economics, 110, 605-639.*

[15] Gilboa, I., Schmeidler, D. (2001). A Theory of Case-Based Decisions. *Cambridge University Press, Cambridge.*

[16] Goldstone, R. (1994). The role of similarity in categorization: Providing a groundwork. *Cognition, 52-2, 125-157.*

[17] Hayne, H. (1996). Categorization in Infancy. *in Advances in Infancy Research, 10.*

[18] Jones, G. Y. (1983). Identifying basic categories. *Psychological Bulletin, 94, 423-428.*

[19] Krueger, J., Clement, R. (1994). Memory-based judgments about multiple categories. *Journal of Personality and Social Psychology, 67, 35-47.*

[20] Kruschke, J.K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99, 22-44.*

[21] Malt, B. C., Ross, B. H., Murphy, G. L. (1995). Predicting features for members of natural categories when categorization is uncertain. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21, 646-661.*

[22] Manzini, P., Mariotti, M. (2012). Categorize Then Choose: Boundedly Rational Choice and Welfare. *Journal of the European Economic Association, 10-5, 1141-1165.*

[23] Medin, D. L., Schaffer, M.M. (1978). Context theory of classification learning. *Psychological Review, 85, 207-238.*

[24] Medin, D. L. (1983). Structural principles of categorization. *in Interaction: Perception, Development and Cognition, 203-230.*

[25] Medin, D. L., Aguilar, C. (1999). Categorization. *in MIT Encyclopedia of the Cognitive Sciences.*

[26] Mohlin, E. (2014). Optimal Categorization. *Journal of Economic Theory, 152, 356-381.*

[27] Mullainathan, S. (2002). Thinking Through Categories. *MIT Working Paper.*

[28] Murphy, G. L., Medin, D.L. (1985). The Role of Theories in Conceptual Coherence. *Psychological Review, 92, 289-316.*

[29] Murphy, G. L., Ross, B. H. (1994). Predictions from uncertain categorizations. *Cognitive Psychology 27, 148-193.*

[30] Murphy, G. L. (2002). The Big Book of Concepts. *MIT Press, Cambridge.*

[31] Osherson, D. N., Wilkie, O., Smith, E. E., Lopez, A., Shafir, E. (1990). Category-Based Induction. *Psychological Review, 97, 185-200.*

[32] Peski, M. (2011). Prior symmetry, similarity-based reasoning, and endogenous categorization. *Journal of Economic Theory 146, 111-140.*

[33] Posner, M. I., Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology, 77, 353-363.*

[34] Pothos, E. M., Chater, N. (2002). A simplicity categorization. *Cognitive Science, 26, 303-343.*

[35] Quinn, P.C., Eimas, P.D. (1996). Perceptual Organization and Categorization. *in Young Infants: Advances in Infancy Research, 10, 1-36.*

[36] Reed, S. K. (1972). Pattern Recognition and Categorization. *Cognitive Psychology, 3, 382-407.*

[37] Rips, L. J. (1975). Inductive Judgements About Natural Categories. *Journal of Verbal learning and Verbal Behavior, 14, 665-681.*

[38] Rips, L. (1989). Similarity, typicality, and categorization. *in Similarity and analogical reasoning, 21-59.*

[39] Rosch, E., Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7, 573-603.*

[40] Rosch, E., Lloyd, B. (1976). Cognition and categorization. *Lawrence Erlbaum Associates, Hillsdale.*

[41] Rosch, E. (1978). Principles of Categorization. *in Cognition and Categorization, 27-48.*

[42] Savage, L.J. (1954). The foundations of statistics. *New York, John Wiley and Sons.*

[43] Smith, L. B. (1989). From global similarity to kinds of similarity: The construction of dimensions in development. *in Similarity and analogical reasoning, 146-178.*

[44] Smith, E.E., Patalano, A.L., Jonides, J. (1998). Alternative strategies of categorization. *Cognition, 65, 167-96.*

[45] Tversky, A. (1977). Features of Similarity. *Psychological Review, 84, 327-352.*

[46] Vanpaemel, W., Storms, G. (2008). In search of abstraction: The varying abstraction model of categorization. *Psychonomic Bulletin and Review, 15-4, 732-749.*