Center for
Mathematical Economics
Working Papers

507

April 2014

Cautious Belief Formation Model

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

Cautious Belief Formation

Jörg Bleile*

April 15, 2014

Abstract

We provide an axiomatic approach to a belief formation process in an informational environment characterized by limited, heterogenous and differently precise information. For a list of previously observed cases an agent needs to express her belief by assigning probabilities to possible outcomes. Different numbers of observations of a particular case give rise to varying precision levels associated to the pieces of information. Different precise information affects the cautiousness and confidence with which agents form estimations. We modify the Concatenation axiom introduced in Billot, Gilboa, Samet and Schmeidler (BGSS) (Econometrica, 2005) in a way to capture the impact of precision and its related perceptional effects, while still keeping its normative appealing spirit. We obtain a representation of a belief as a weighted sum of estimates induced by past cases. The estimates are affected by cautiousness and confidence considerations depending on the precision of the underlying observed information, which generalizes BGSS. The weights are determined by frequencies of the observed cases and their similarities with the problem under consideration.

Keywords: Belief formation, prior, relative frequencies, case-based reasoning, similarity precision, ambiguity, learning, imagination, confidence, cautiousness.

JEL Classification: D01, D81, D83

*Center for Mathematical Economics (IMW) and Bielefeld Graduate School of Economics and Management (BiGSEM), Bielefeld University, jbleile@uni-bielefeld.de.
I would like to thank Raphaël Giraud, Christoph Kuzmics and Frank Riedel, for helpful comments and suggestions.
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 and Motivation

Beliefs of agents are important ingredients in many economic models dealing with uncertainties. Belief formation is studied recently in environments with limited and heterogenous information, that are not suitable to be modeled in the widely used and accepted state space framework of Savage (1954) and Bayes. Lacking a state space representation of uncertainties an agent needs to form her belief explicitly by directly using available information.

We axiomatize a belief formation process based on limited, differently relevant and precise available information. Our main axiom modifies the concatenation axiom in Billot et al. (2005)(BGSS), which precludes the impact of agents' perceptions and reactions to differently precise information. Their axiom says, that for any two information sets the belief induced by their combination can be expressed as a weighted average of the beliefs induced by each information set separately. The averaging of beliefs induced by any arbitrary information sets requires a cognitive challenging tradeoff of identical, but differently precise information contained in the particular information sets. Our axiom says, that agents, that care about precision of information, can only average beliefs (in a normatively reasonable way) induced by specific - almost disjoint - information sets. Thereby, we focus not only on the precision itself, but also on its perception and impact in form of cautiousness and confidence feelings.

The most prominent and often used models to describe and analyze uncertainty in economic theory are versions of the approach of Savage and Bayes. The fundamental idea in this approach is to model uncertainty by a grand state space, which is sufficiently rich to describe and resolve all possible sources of uncertainties. In this way a state space implicitly incorporates some (perfect) belief (and theorization about structures and relationships) about the future and thereby requires a large (often un-achievable) imagination and theorization task of agents. In addition, insufficient (or too complex) information may preclude the derivation or definition of a grand state space. Another principle of the state-based approach is the representation of a belief as an unique probability over the grand state space. In this framework a purely subjective probability distribution over states can be endogenously deduced from preferences, which inherently lacks an explicit description of the formation of the belief that generated the preference [1].

There is basically only one way to deal with these two difficulties. Sticking to the grand state space-principle, but abandoning the subjective prior approach, would precludes an direct (objective) assignment of probabilities, since the state space already encodes all available information. More promising is to give up the representation of uncertainty by a state space, when an agent is (cognitive) incapable to translate information into (imagined and theorized) states or the information is not all encompassing as needed for the "correct" description of a grand state space. In many real-life situations list of pieces of information (databases) represent our informational basis. We will replace the state space as an information aggregation by such a database representation of (actually observed) information (data-points or recalled cases).

A belief based on a database needs to explicitly incorporate factual objective knowledge,

[1] See Gilboa et al. (2012) for extensive discussion of these issues.

characteristics and theoretical considerations provided by the present database. In general, belief formation based on a database is very close to the goal of statistical inference. In contrast to mainly asymptotic considerations in statistical inference, our focus (as in BGSS and Eichberger and Guerdjikova (2010) (EG)) lies on behavioral foundations (axiomatizations) of a belief formation and the analysis of small databases containing differently precise information.

Usually statistical experiments are dealing with identical observations, which are equally relevant. However, since small or medium sized databases contain limited and heterogenous information agents might want to take into account not only (a few) identical but also partially relevant observations for their belief formation. In this sense - differently to statistical experiments - relevance or similarity measures become important, when data sets contain limited heterogenous information.

Case-based Decision Theory (Gilboa and Schmeidler (1995, 2001)) deals with such a framework in decision theoretic contexts. BGSS can be interpreted as an adoption of it to belief formation [2]. Their axiomatized belief describes a generalized (subjective) frequentist, in which agents assign different similarity weights to information with different degree of relevance. For a new problem and given a database of past observations, their belief over possible outcomes is represented as a similarity weighted average of estimates, that are induced by the observed cases.

Their main concatenation axiom deals with relationships between databases and their induced belief. It requires that for a new problem x the belief P (probability vector over outcomes for x) induced by the combination (concatenation) of any two databases ($D \circ E$), is a weighted average of the belief induced by each single database (D and resp. E) separately, i.e. for all databases D and E, there exists a $\lambda \in (0, 1)$ such that

$$P(x, D \circ E) = \lambda P(x, D) + (1 - \lambda)P(x, E)$$

Our paper deals mainly with the modification of this concatenation axiom in order to allow for impacts of precision of information and its induced cautiousness and confidence concerns. Additionally, our precision dependent belief formation is suitable for small databases, which is only partially possible and reasonable for BGSS.

The concatenation axiom shows some irrelevance of growing precision. The belief induced by a database coincides with the belief induced by arbitrary many replications of the same database, i.e. $P(x, D) = P(x, D^T)$ for all $T \in \mathbb{N}$. Growing precision might not be a concern for sufficiently rich and large databases such that observing additional identical information will not affect her predictions. However for small database, specifically consisting only of one piece of information c, it is unreasonable that additional observations do not induce some learning and refinement of an already "perfect" estimation, i.e. $P(x, c^T) = P(x, c)$ for all T.

In this way the concatenation axiom implies that one observation carries already anything, that can be inferred by arbitrary many confirming observations. Such a instantaneous learning in a highly objective (and perfect) way of forecasting appears to be questionable and un-intuitive. For instance consider a situation, in which an agent throws a

[2] Related also to Gilboa and Schmeidler (2003) and Gilboa et al. (2011) and related Gilboa (2009).

dice once and the figure six results. A guess of the outcome of the next throw of the dice would differ form the estimation an agent would come up after observing one million times a six in one million throws of that dice. However roughly speaking, the concatenation axiom requires that an agent would infer right after the first dice throw that all sides of the dice show the figure six, without any doubt. A procedure to base the estimation on just one observation appears to be in-cautious, hazardous (error-prone) and unrealistic and cannot be considered as an appealing normative advice. In fact, as in controlled statistical experiments, additional (identical) confirming observations may serve as a proxy for its increased informativeness, precision or accuracy, which should be reflected in a dynamic learning and refinement of the estimations.

In addition, increasing precision might affect estimations through its perception in form of altered cautiousness (to wrongly eliminate some outcomes) with which the forecast is made and her changed confidence in this forecast[3]. If information becomes more precise, an agent's decreased cautiousness and increased confidence might allow to specify their prediction. After receiving substantial information of disconfirming evidence that makes some outcomes negligible, agents even might want to eliminate some (not observed) outcomes. More general, differently precise information should lead to different induced beliefs, i.e. $P(x, c^T) \neq P(x, c^L)$ for different $L, T \in \mathbb{N}$, which contradict the concatenation axiom and requires a modification in order to incorporate precision and cautiousness issues.

In general, the concatenation axiom is stated for any kind of databases, but (with regard to potentially induced different precise estimations) it is most appealing and appropriate as a normative advice for disjoint databases. For disjoint databases, the belief induced by the concatenated database can be quite intuitively interpreted as an average of the beliefs induced by the single databases separately, since no pieces of information appear in different precision in different databases and cause conflicting considerations. The average is determined solely by a weighting of the relevances of the concatenating databases.

However, we will explain, that for unrestricted non-disjoint databases - with common, but differently precise pieces of information - the normative appealing spirit of averaging beliefs conflicts with a simultaneous care about precision and cautiousness in the belief formation [4].

A first obvious modification deals with the issue, that a precision related concatenation axiom cannot be formulated for pure (non-disjoint) concatenating database, but would require additional information, such that averaging occurs according to $P(D \circ E) = \lambda P(D^L) + (1-\lambda) P(E^T)$ for some appropriate $L, T \in \mathbb{N}$ (which we will specify later). For example consider the concatenation of the easiest non-disjoint databases $(c^7) = (c^3) \circ (c^4)$.

By definition, the beliefs induced by combining databases (e.g. $(c^3), (c^4)$) are based on less precise pieces of information (and hence also their weighted average) than the beliefs based on the combined database (e.g. (c^7)). Thus, for a cautious agent, that cares about precisions, the information contained in concatenating databases may not be sufficient for a belief formation according to the (unmodified) concatenation axiom[5].

[3]See Ellsberg (1961) (p.657): "What is at issue might be called the ambiguity of this information, a quality depending on the amount, type, reliability, and "unanimity" of information, and giving rise to ones degree of "confidence" in an estimate of relative likelihoods."

[4]Stating the axiom only in terms of disjoint databases does not offer sufficient structure to derive a belief formation.

[5]This problematic issue does not appear in the concatenation axiom of BGSS, where precision is (endogenously) neglected and one appearance of a case captures already all information.

However, in general there are no replications T and L, such that each single pieces of information is captured in equal precision in all involved databases $D \circ E$, D^T and E^L. These differences in precision of single common cases complicate the averaging of the beliefs. Determining the average weight cannot anymore be interpreted as normatively appealing comparison based solely on relative relevance of the particular databases. Rather, it is a result of a cognitive challenging (impossible) interwoven tradeoff, balancing and aggregation of different emphasis an agent assigns to single pieces of differently precise information in the various database. Moreover, the average weight might need to reflect also the compensation for failures of the compulsory (by the axiom) incorporation of relative more imprecise estimations (based on the same kind of information) contained in some beliefs than in others. Therefore, a (modified) concatenation axiom allowing for all (replicated) databases leads to the serious problem, that agents might be cognitively overstrained by averaging beliefs based on several identical information with different precision levels[6].

As a consequence, we propose a restriction on databases to be admissible for our modified version of the concatenation axiom, such that it sustains its normative appealing spirit. Our anchored concatenation axiom restricts databases to a specific (almost disjoint) structure consisting of only two cases, where only exactly one of these cases (the anchor) appears in all involved databases. The main feature is, that this single common (non-disjoint anchor) piece of information is contained in all involved databases in equal precision. This enables an easy averaging of beliefs without cognitively demanding compromising between estimations induced by differently precise observations. In addition, the equal appearance of the anchor case in all involved databases intuitively allows to "neglect" its effect in determining the average weights and to compare only the relative importance of the pieces of information, that appear only in exactly one of the involved databases. This facilitates a very straightforward way to find the average weights for the beliefs - almost in the spirit of averaging beliefs induced by disjoint databases.

In order to take into account the precision of beliefs induced by databases, our agent focusses on the most precise and hence reliable information in the database. Since it is impossible to capture all information in its actual precision level perfectly in a non-disjoint combination of databases (as explained above), our agents require to cover at least the most precise information objectively in her belief. Consequently our axiom requires, that there exists no distortion of the most reliable information in the process of averaging beliefs. To achieve this, the precision of most reliable information in the concatenated database must be conveyed by the single beliefs induced by the corresponding admissible (sufficiently replicated) databases used for the concatenation.

Besides BGSS, the closest work to ours is the axiomatization of multi-prior beliefs in Eichberger and Guerdjikova (2010) (EG) "Case based belief formation under ambiguity". Their extension of the framework of BGSS aims to formalize two kinds of ambiguity caused by insufficient information (vanishing ambiguity) and irrelevant information (persistent ambiguity). The focus in their paper lies predominantly on the introduction of a multi-prior setup for an information environment with persistent ambiguity. Whereas our work

[6] Alternatively, if one would stick to general non-disjoint databases, then the only way to "unify" the differently precise information in all involved databases is given by assuming ad hoc some arbitrary (imagined) level of precision, according to which all cases are evaluated independent of their actual observation. This will be discussed in detail in Section 6.1.

focuses on the analysis of precision in the sense of vanishing ambiguity (imprecision) and related cautiousness in a single prior belief. EG's modification of the concatenation axiom of BGSS is adequate to specify how beliefs over outcomes change in response to additional information and tackles also the mentioned drawback of BGSS regarding irrelevance of growing precision of information. Different to our work, their modification reflects the idea of "controlling for the ambiguity " (p.4) (precision) by restricting the involved databases to equal length. However, as discussed above controlling for precision by equal lengths of the involved databases is not sufficient to control for different precise information contained in these databases. As a direct consequence, EG's modification of the concatenation axiom assumes (and does not prevent) that agents are (cognitively) able to aggregate and balance information of the same kind, but in different precision. In contrast, the focus of our paper lies exactly on the issue to avoid such cognitive challenging or even impossible tradeoffs in the aggregation of differently precise information and to keep the spirit of a normatively reasonable and easy averaging procedure. Moreover, in general EG's axiom implies that no estimation is based on objectively present information in the database, which would require (in our context) that agents need to imagine the (true) cautiousness feeling evoked by a precision level that is imagined as well. In contrast our approach implicitly requires only the ability to estimate based on already experienced cautiousness and thus avoids imaginations of unexperienced feelings of cautiousness.

In sum, adopting parts of the axiomatization of BGSS and EG, our anchored concatenation axiom will allow for a structural similar belief representation as in BGSS and EG. All three axiomatized representations differ in the way, how they treat information of different qualities of precision. BGSS does not take into account precision at all and EG captures the effect of (persistent) ambiguous information by a set of beliefs, which are based on precision-dependent estimates, where the level of (imagined) precision is according to the total amount of information contained in the entire database. In our representation the cautiousness related estimates are based on the level of precision and cautiousness induced by the most precise information in the database. More precisely, for a new problem and a given database, its induced belief can be represented as a similarity-weighted average of cautious estimates induced by past observations in the database. Thereby the similarities and estimates are endogenously derived.

The remainder of the paper is organized as follows. In the next section we will outline the model and in Section 3 we develop an example to illustrate reasonable belief formations and our leading example. Then the axioms are stated and discussed, where the central Section 5 points out the drawbacks and necessary modification of the concatenation axiom to incorporate precision, which eventually leads to our version of the concatenation axiom. Section 6 presents and discusses the main representation result. Section 7 concludes the paper. Appendix A and B contain both directions of the proof, where B.4 gives a rough sketch of the main part of the proof. The rest of the Appendix deals with an objective belief, the relationship to EG's axiom and an alternative axiomatization of a very cautious belief.

2 The model

2.1 Cases and Databases

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, $R = \{r^1, ..., r^m\}$.
The set $C \subseteq X \times R$ consists of all m basic cases, i.e. $|C| = m$.
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. $D = ((x_1, r_1), ..., (x_L, r_L))$ is denoted by C^L and the set of all databases by $C^* = \cup_{L \geq 1} C^L$. 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.
We need some definitions for the framework of database.
The concatenation of two databases $D = (c_1, c_2, ..., c_L) \in C^L$ and $E = (c'_1, c'_2, ..., c'_T) \in C^T$ 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.
If a case $c \in C$ appears in a database D, i.e. $f_D(c) > 0$, we write $c \in D$.
Two databases D and E are called disjoint if for all $c \in C$: $c \in D$ if and only if $c \notin E$.

2.2 Induced Beliefs

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 $\{1, 2, ..., n\}$.
An agent will form a belief over the 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.
$P : X \times C^* \rightarrow \Delta(R)$. The restriction to databases of length T is denoted by $P_T(x, D) \in \Delta(R)$ for $D \in C^T$ and $P_T : X \times C^T \rightarrow \Delta(R)$.
One can interpret $P_T(x, D)$ as the belief over outcomes induced by database $D \in C^T$ (given environment or problem $x \in X$).
Throughout the paper the problem x is fixed, therefore x is often suppressed in the following, i.e. $P(x, D) = P(D)$.

3 Motivating examples

3.1 Exemplary development of a belief formation process

A doctor needs to evaluate the likelihood of potential outcomes of a specific treatment.
Let a patient be described by a vector of characteristics $x \in X$, where X might consist of measures of characteristics like age, gender, weight, height, blood pressure, temperature, blood count, vital signs, medical history, drug tolerability, etc.
The doctor might have observed several outcomes of the treatment in the past, which are collected in a set R, containing e.g. feels better, worse or unchanged, measures of side

effects like headaches, sleepy, depressive, passed out, giddy, dizzy, etc.
The doctor has acquired some working experience prescribing this treatment and/or has access to some medical record on this treatment. Thus, she is able to base her judgement on past experience or observations collected in a database $D = (c_1, ..., c_T)$, where in each case c_i the characteristic and the observable outcome of patient i is recorded, i.e. $c_i = (x_i, r_i)$, where $x_i \in X$ and $r_i \in R$. It means that a patient characterized by x_i responded to the treatment with outcome r_i.
Given the characteristics $x \in X$ of a current patient and her available information and experience in form of a database D, the doctor derives a probabilistic belief $P(x, D) \in \Delta(R)$ over potential outcomes in R for this treatment. How can she do the evaluation?

a) A first intuitive approach for the prediction is, to consider only patients in the database, which are identical (with respect to the measured characteristics) to the present patient. Based on this sub-sample $D_x := (c \in D | c = (x, r_i) \text{ for some } r_i \in R) \subseteq D$ the doctor might derive a prediction over potential outcomes via **empirical frequencies**:

$$P(x, D) = \frac{\sum_{c_j \in D_x} \delta_{r_j}}{|D_x|},$$

where δ_j is the probability vector on R with mass 1 on the outcome $r_j \in R$. Of course this belief formation process is not practical, if the sub-sample D_x contains only few observations, i.e. if there are only a couple of identical (with respect to the measured characteristics) patients.

b) To overcome this problem of limited or insufficiently many identical observations, the doctor might include into her prediction procedure not only identical, but in addition also similar patients. Suppose, that she is able to judge how similar patients are, i.e. she is able to employ a function s : $X \times X \to \mathbb{R}$, where $s(x_t, x_j)$ measures the degree of similarity between patients characterized by x_t and x_j. Her belief formation process might run in a **"subjective" frequentist** way:

$$P(x, D) = \frac{\sum_{c_j \in D} s(x, x_j)\delta_{r_j}}{\sum_{c_j \in D} s(x, x_j)}$$

c) In addition, the doctor might infer from a case $c_j = (x_j, r_j) \in X \times R$ not only a point prediction on δ_{r_j}, but a more general induced estimation P^{c_j} [7]. Basically, she attaches also some likelihood to outcomes that are closely or reasonably related to the observed r_j:

$$P(x, D) = \frac{\sum_{c \in D} s(x, x_j) P^c}{\sum_{c \in D} s(x, x_j)}. \tag{1}$$

This belief formation process is axiomatized in **BGSS**.

d) Furthermore, the doctor might process the past observations not in an one by one estimation problem as in the approaches above, but might want to sample the database

[7] More precisely, actually $P^c = P^{(x,c)}$ represents an estimate induced by c given the current patient x, i.e. if c is totally unrelated to the current patient, it might be that P^c is uniform on R.

beforehand according to identical cases. Many observations of the same case might foster some learning and improved understanding of the relationship between characteristics of a patient and corresponding outcomes. Additional confirming observations should affect the judgment of a cautious doctor as well by an increased confidence and decreased cautiousness in predicting the observed outcome. In this way, the doctor might generate different predictions depending on how many observations of this case are present in the database.

For instance, suppose there is a generally observed side-effect of many different medicines, then the doctor might still assign a positive likelihood to this side-effect, if the doctor has observed just a few (identical) patients not suffering from this side effect under the specific treatment. However, if the treatment is well established and many identical patients did not feel this side-effect, then she might not consider this side-effect as a potential hazard anymore. This intuition can be modeled by incorporating precision into a cautious belief formation, where the number of observation can be interpreted as a proxy for the precision of the information:

$$P(x,D) = \frac{\sum_{c\in D} s(x,c) f_D(c) P^c_{T_D(c)}}{\sum_{c\in D} s(x,c) f_D(c)},$$

where P^c_L represents the precision dependent estimation on R induced by L observation of case c, where usually $P^c_T \neq P^c_L$ for $T \neq L$. Hence $T_D(c) \in \mathbb{N}$ denotes a database dependent precision (and induced cautiousness) level, according to which a doctor will estimate the outcomes based on observation of case c.

Interestingly, the already mentioned belief formations of BGSS and EG are special cases of this representation:
(i) BGSS's axiomatization implies $T_D(c) = \infty$. Hence, their agent learns instantaneously the "correct" distribution $P^c_\infty = P^c$ induced by case c (see representation (1)).
(ii) The axiomatic derivation in EG results in $T_D(c) = T$ for $D \in C^T$.

(e) A natural interpretation of $T_D(c)$ in (d) is $T_D(c) = f_D(c)T$ for all $c \in D \in C^T$:

$$P_T(x,D) = \frac{\sum_{c\in D} s(x,c) f_D(c) P^c_{f_D(c)T}}{\sum_{c\in D} s(x,c) f_D(c)}, \tag{2}$$

where $f_D(c)T$ gives the actual number of appearance of case c in database D. Such a representation is very objective by incorporating only actually available and observed information. This representation is (unfortunately) irreconcilable (see Appendix C) with any generalized version of a concatenation axiom (in the sense of not only combining disjoint databases), which is an important behavioral component of a belief formation.

However, unless its appealing objective character, this belief formation might entail the following problem. Obviously, the belief employs (in general) an aggregation of estimates $(P^c_{f_D(c)T})_{c\in D}$ based on different precise single pieces of information c, which carry different deficits in their correctness of prediction. This might over-complicate the evaluation since the doctor might want to accompany the fact, that some of her predictions are more reliable than others and should receive more weight independent from the similarity and frequency weighting. In this sense, she might want to include an additional weighting scheme taking into account the precision or reliability of the estimates [8].

[8] Alternatively, these considerations might be incorporated into the weights s, which prevents an desirable independent

f) However, a doctor might not only rely on objective precisions of the estimations, but also wants to capture her perception of its precision, i.e. the influence of how cautious and confident she feels while estimating $P^c_{T_D(c)}$. In this vein, we prefer a different choice for $T_D(c)$ in order to take the doctor's cautiousness and confidence concerns into account.

The underlying intuition is that she does not change or adjust constantly her cautiousness and confidence attitude in response to each differently precise information. Rather, after the doctor has experienced an (extreme) level of cautiousness and confidence by estimating based on objectively available (unimagined) information, she might keep and adopt it to other estimations. Basically she attained an "appropriate" sustainable attitude regarding her cautiousness sensation or learned how to confidently estimate sufficiently cautious and applies it to all remaining estimations. This also overcomes the mentioned potential disfavor of aggregating different precise estimations emerging in the objective belief formation (2).

The most intuitive choices for a cautiousness attitude are the two extreme perceptions, i.e. the experience of minimal and maximal cautiousness, which are induced by the most or least precise information in the database. A minimal cautious attitude might distract from any other more cautious perceptions, since the doctor learned how to handle information in an appropriate cautious way. A maximal cautious agent might be intimidated by the experienced imprecision and can not be convinced to leave her skeptical mood to adopt a more confident attitude for estimating according to the available more accurate information.

The following **cautious belief formation** captures these ideas (for a attitude of minimal cautiousness) and will be axiomatized in our paper:

$$P_T(x, D) = \frac{\sum_{c\in D} s(x,c) f_D(c) P^c_{\max_c f_D(c)T}}{\sum_{c\in D} s(x,c) f_D(c)}.$$

The above examples were intended to clarify the framework and demonstrate a meaningful evolution of a belief formation taking into account subjective and precision concerns. However, in the following we will use a reduced version as our leading example.

3.2 Leading Example

Assume that, the patients are not anymore described by a large vector of their personal characteristics, but just according to their symptoms or diagnosed sickness. In particular, each patient is characterized by just a single symptom and the outcome of a treatment is only roughly distinguishable between w(orse), n(ot affected) or b(etter), i.e. $R = \{w, n, b\}$.

So basically a doctor has prescribed a certain medicine to many patients with different symptoms or illnesses and observed the outcome of this drug, i.e. a case is described as a pair of symptom and outcome of the treatment. For example the drug improved on the state of patients suffering from sore throat, but was harmful for most patients suffering from stomach problems.

interpretation in similarity terms and requires the function s to depend also on the databases directly, which will be precluded (later) by our constant similarity axiom. In addition, it conflicts with the easy averaging intuition

4 Axioms

In the first part we adopt modified versions of the uncritical axioms in BGSS. The second main part discusses in detail the concatenation axiom and its drawbacks in a precision dependent framework, which eventually leads to our new anchored concatenation axiom.

4.1 Uncritical Axioms

4.1.1 Invariance Axiom

For every $T \geq 1$, every database $D = (c_1, ..., c_T) \in C^T$ and any permutation π on $\{1, ..., T\}$

$$P_T((c_1, ..., c_T)) = P_T((c_{\pi(1)}, ..., c_{\pi(T)}))$$

The Invariance axiom states, that an induced belief over outcomes depends only on the content of that database and is insensitive to the sequence or order in which data arrives.

However, the order in which information is provided or obtained can influence the judgment strongly and may carry information by itself (e.g. see Rubinstein and Salant (2006)). For example, first and last impressions or reference effect demonstrate the different impacts of cases depending on their positions. 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 a database, one could implement this information into the description of the cases itself. Put differently, if one challenges the invariance axiom, then there must be some criteria which distinguish the cases at different positions in a database and paying attention explicitly to this difference in the description of the cases may lead the agent to reconcile with the invariance assumption.
Hence, we will base our belief formation only on the content of the database D, which allows to characterize each D by the pair of its frequency vector and length, i.e. $(f_D, |D|)$.

4.1.2 Learning Axiom

For every $c \in C$ the limit of $P_T(c^T)$ exists, i.e. the sequence converges to P_∞^c.

In the context of precision dependent beliefs the axiom can be interpreted as a stable learning process. For instance, an agent starts out with an initial prior (like a uniform as in the principle of insufficient reason) that will be adjusted in the process of observing additional information. Increasing the number of confirming observations will lead to vanishing imprecision and cautiousness in estimating. Basically, the estimate will become less sensitive to new additional confirming information and will eventually converge to a limit distribution. This intuition is as in Bayesian updating, where additional (confirming) information may render the prior beliefs more precise, but differently to Bayesian updating the support might change here. For instance, it is reasonable to assume that finally the agent will learn the true distribution of a case $c = (x, r) \in C$ given the problem $x \in X$, i.e. $\lim_{T\to\infty} P(x, c^T) = \delta_r$, where δ_r is again the Dirac measure. However for a problem $x' \neq x \in X$ the belief might just converge to a general uniform-like distribution on R, since the observed case does not give relevant information for the current problem at all. Hence, we require only that such a limit estimation exists.

Another intuition that we mentioned already, runs as follows. T many observations a case $c = (x, r)$ might not make a cautious agent feel confident to reliably rule out a non-observed outcomes completely, but she wants to assign at least some positive likelihood to it, i.e. $P_T((x, r)^T)(r') > 0$. However, observing further confirming cases might carry sufficient evidence, such that an agent would feel confident not to make a mistake or act incautious in excluding some outcomes, i.e. $P_L(x, r)^L)(r') = 0$ for $L \gg T$.

Alternatively, one can apply a (accordingly adjusted) learning procedure as in Epstein and Schneider (2007), where an agent might start out with a uniform estimation and after observing new information keep only the most plausible estimates. Plausible estimations in their sense are those that survive a maximum likelihood test (according to some strictness parameter, which might correspond with a cautiousness measure in our setup) against the belief that best explains the observations, i.e. the dirac measure on the observed outcome.

4.1.3 Diversity Axiom

There exist $T^* \in \mathbb{N}$, such that for all $T \geq T^*$, no three of $\{P_T(c^T)\}_{c \in C}$ are collinear.

Form a technical point of view this axiom allows to derive an unique similarity function, but it also carries an appealing intuition. Roughly it states, that sufficiently many observations induce always estimations, which are informative (or diverse) in the sense that no combination of two other sufficiently often observed cases can deliver the same estimation. Hence, no sufficiently precise case can be "replaced" by sufficient observations of two other cases in this sense. The reason to base the diversity of induced estimation on a precision threshold T^* is the following. In order to derive unique similarity values one could also require non-collinearity for every value of T, but this would exclude learning as mentioned in the description of the learning axiom. If an agent would start out with an uniform-like prior for databases containing few observations, it might happen that different cases induce very similar estimations, which are likely to be collinear. The axiom just rules out, that after a sufficient learning period any three estimations are still collinear.

4.2 Different Versions of the Concatenation Axiom

4.2.1 Concatenation Axiom of BGSS

For every database $D, E \in C^*$ there exists some $\lambda \in (0, 1)$ such that:

$$P(D \circ E) = \lambda P(D) + (1 - \lambda)P(E).$$

In the following we will call the database which emerges from concatenations of other databases the **combined or concatenated** database, whereas the databases used for the concatenation will be called **combining or concatenating** databases. We call the weights $\lambda, (1 - \lambda)$ **average weights.**

The concatenation axiom states that the belief induced by a combined database is a weighted average of the beliefs induced by its combining databases. It captures the idea that a belief based on the combined database 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 (from an exclusion point of view): if the information in any database induces an belief that does not exclude an outcome r, then the outcome r cannot be excluded by the belief induced by the combination of all these databases [9]. Alternatively, if a certain conclusion is reached given two databases, the same conclusion should be reached given their union.

The normatively appealing spirit of the axiom is that the average weights are determined by relative relevances or importance of the combining databases for its combination.

As already mentioned, the Concatenation axiom implies an irrelevance of growing precision or insensitivity to additional information in the beliefs, i.e. $P(D) = P(D^Z)$ for all $D \in C^*$ and $Z \in \mathbb{N}$, which might be appropriate for sufficiently rich and large databases. However, already BGSS admit, that it "... might be unreasonable when the entire database is very small ...." (BGSS (2005), p. 1129)[10]. Indeed, the axiom induces some sort of perfect objectivity and instantaneously learning. Estimation based on one observation $c = (x, r)$ needs to coincide with the estimation induced by arbitrarily many observation, which can be identified in some sense with the "true" limiting distribution, i.e. $P(x, (x, r)) = P(x, (x, r)^\infty)$. For our leading example it would mean that a doctor would predict after one unsuccessful treatment of a sore throat that this treatment is worthless for (identical) patients suffering from sore throat. However, this appears very unrealistic and un-intuitive, since a database $c = (x, r)$ might be considered more imprecise and might induce a more cautious belief than $c^T = (x.r)^T$ for sufficiently many observations T, i.e. $P(c) \neq P(c^T)$.

In order to incorporate precision and cautiousness aspects into the belief formation process, the concatenation axiom needs to be modified in various ways to maintain its normative appeal in a modified framework.

For this purpose an immediate modification concerns the issue that an agents can not rely on beliefs induced by the concatenating databases directly, but requires appropriately replicated concatenating database [11], i.e.

$$P(D \circ E) = \lambda P(D^T) + (1 - \lambda) P(E^L) \text{ for appropriate } T, L \in \mathbb{N} \tag{3}$$

The reason for that is that the information contained in non-disjoint concatenating databases appears by definition in less precision as in their concatenation, e.g. consider $c^{2Z} = c^Z \circ c^Z$. However, for a cautious agent caring about precision, $P(c^Z)(r) > 0$ does not necessarily imply $P(c^{2Z})(r) > 0$. For example our doctor might not want to rule out a successful treatment of a coughing agent after observing 20 or 30 unsuccessful treatments according to her perceived cautiousness, i.e. (i.e. $P(c^T)(r) > 0$ for $T = 20, 30$). However, the combined information of 50 unsuccessful treatments on coughs might make her feel confident and convinced to evaluate the treatment as useless for curing a cough without violating her cautiousness feeling, i.e. $P(c^{50})(r) = 0$. Thus, non-disjoint concatenating databases

[9] 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 databases.

[10] From this perspective, our modification can be interpreted as an extension of BGSS to derive a belief formation also for relatively small databases, which is only partially possible and reasonable given their concatenation axiom.

[11] This problem emerges only if the databases are non-disjoint. However to allow only disjoint databases in the concatenation axiom does not offer enough structure to derive a belief.

do not carry sufficient information to capture refinements of a cautious belief implied by the concatenated database and as stated in (3) more precise (i.e. appropriately replicated) concatenating information is required.

Furthermore, in general there exist no replications T and L that ensure that each case in $D \circ E$ is captured in identical precision for unrestricted non-disjoint D^T and E^L [12]. Depending on its precision the same case might induce differently cautious estimations. This leads to the difficulties that agents need to balance the differently cautious estimations induced by the same case appearing in different precisions in the replicated concatenating databases. Such a compromising between estimates is necessary for all cases, that are observed in more than one database. For instance, our doctor compares the (replicated) databases $D^2 = (c_1^4, c_2^6, c_3^4)$ and $E^2 = (c_1^4, c_2^8, c_3^2)$ (and eventually average its induced beliefs), where each (replicated) database contains differently many observations of harmfully treated colds c_2 (6 vs. 8), neutrally treated colds c_3 (4 vs. 2) and at least the successfully treated sore throats c_1 are observed identically often (4 vs. 4) (by replicating D and E, with the focus on unifying according to c_1). Hence each induced beliefs rely on different precision with regard to observations of cases c_2 and c_3. How could an objective doctor compare and average the differently precise information incorporated in these databases? Intuitively, the doctor should use the most precise available information contained in these databases. Information c_2 is contained in the belief induced by E^2 in a more precise fashion than in database D^2 and hence the doctor would like to rely predominately on (i.e. assign high weight to) E^2 regarding c_2 (since $P_8^{c_2}$ vs $P_6^{c_2}$) and to ignore the less precise estimation wrt. c_2 in D^2. However, the opposite is true for the precision of information c_3, for which she relies predominately on D^2 and ignores E^2.

However, such a reasonable behavior is not admissible in any version of a concatenation axiom, where an agent is forced to assign exactly one (non-zero) average weight to the beliefs induced by the entire databases D^2 and E^2 and not many different weights to the estimates induced by the single pieces of information contained in the databases [13]. In order to reach one "aggregated" average weight, these single weights would need to be balanced, traded off and aggregated somehow. In particular, since the beliefs induced by D^2 and E^2 contain induced estimates, that are too imprecise and cautious in comparison to other available ones, our doctor needs to offset and capture these imprecisions and mistakes by adjusting the average weights accordingly. However, a determination of the average weight as a result of difficult balancing and interwoven compromising appears to be even in this easy example rather cognitively challenging and becomes impossible for more complex (decompositions of) databases. Further and most importantly, it conflicts with the normatively appealing spirit of the concatenation axiom to average beliefs by an easy comparison of relevances of the particular underlying databases.

Our modification of the concatenation axiom will deal with this problem by restricting it to specifically structured database such that balancing and compromising due to differently precise information is avoided and the cognitively simple averaging intuition sustains.

[12] This follows from the general non-existence of solutions T, L to the system of equations resulting from $f_D(c)|D|T = \{0, f_{D\circ E}(c)|D \circ E|\}$ and $f_E(c)|E|L = \{0, f_{D\circ E}(c)|D \circ E|\}$ for all $c \in D \circ E$.

[13] However, this potential cognitive difficulties are not an issue in the way the concatenation axiom of BGSS processes information, where information is additive in the sense that L observations in one database and T observations in another is equivalent to observing $T + L$. Since one observation caries all information, literally only the entire amount of appearance is important for the average weight.

However interestingly, following the idea of a concatenation axiom that still allows for unrestricted non-disjoint concatenating databases would eventually arrive at an (not yet given) intuition and explanation for the modification of the concatenation axiom followed in EG. The basic idea is that agents tackle the immense compromising considerations of different cautious estimations by assuming or choosing a common arbitrary level of precision, according to which all cases are estimated - independent of their true objective precision. Since objective or imagined precisions might evoke different feelings of cautiousness such an approach would interfere with our purpose to seriously take into account objective precision and its related concerns. A more detailed discussion on that and on EG's modification can be found in the Appendix D.

4.2.2 Anchored Concatenation Axiom

The above discussion shows that a concatenation axiom for unrestricted non-disjoint concatenating databases might destroy the underlying normatively appealing idea of an easy averaging, when agents care about precision and its perceptional consequences. In order to keep the normative appealing spirit, we will restrict the involved databases to a specific reasonable structure. These databases will contain sufficiently precise information (in the sense of (3)) and allow an cognitively easy averaging. We have seen that an agent will run into a difficult balancing process to determine the average weights when she is faced with concatenating databases containing common cases. For this reason, our anchored databases are as disjoint as possible, but still sharing a specific (exploitable) structure, to facilitate an easy comparison (and in the end a straightforward averaging of its induced beliefs). In particular, the anchored databases consist of only two different cases, where all anchored databases admissible for the concatenation contain a common anchor (reference) case with identical frequency and one additional mutually different case in each of the databases[14]. Besides the desire to employ databases that are almost disjoint, their structure is also driven by the general observation, that agents can compare items easier, if they consist of less features (here: only two) and if they contain common features in the same fashion as a reference (here: anchor case).
Recall, that $m \in \mathbb{N}$ denotes the number of basic cases, i.e. $|C| = m$.

Definition 4.1
Let $k \in [0,1]$ and $T_j \in \mathbb{N}$ be s.th. $kT_j \in \mathbb{N}$ for all $1 \leq j \leq m$ and let $T := \sum_{j \neq i \leq m} T_j$. Let $c_i, c_j \in C$ for all $j \neq i \leq m$
(i) For all $j \neq i \leq m$ a database $D_i^j(k, T_j) \in C^{T_j}$ defined by

$$D_i^j(k, T_j) : \;=\; (c_j^{(1-k)T_j}, c_i^{kT_j})$$

*is called an **anchored database** of length T_j with **non-anchor case** c_j (for all $j \neq i$) and **anchor (case)** c_i, which appears in the database with frequency k.*
*(ii) An **anchored chain** $F \in C^T$ (wrt. to case c_i) is defined as a concatenation of anchored*

[14] In some sense, one can interpret the restriction to such database by agents feeling to only being cognitively skilled or capable to confidently compare such easily structured databases.

databases $D_i^j(k,T_j) \in C^{T_j}$ *for all* $j \neq i$ *(with common anchor case* $c_i \in C$*), i.e.*

$$F = \circ_{j\neq i\leq m} D_i^j(k,T_j) = (c_1^{(1-k)T_1}, ..., c_{i-1}^{(1-k)T_{i-1}}, c_i^{kT}, c_{i+1}^{(1-k)T_{i+1}}, ..., c_m^{(1-k)T_m})$$

Note, that not all databases can be interpreted as an anchored chain, since it requires to be a result of a concatenation of specifically structured anchored databases.

In order to illustrate the anchor-framework, we use our leading example of a doctor, that forms a belief over the outcomes of a treatment -worse, no effect, better- $\{w, n, b\}$. For our doctor anchored databases and chains might look as follows. Each involved (anchored) database consist of only two different cases (patient groups), where one of these groups (the anchor case) needs to be observed in all involved database, e.g. patients with a successful treatment (b) of their cough (c) might be the anchor group (i.e. $c_1 = (c, b)$). The other patient group observed in each database is different in all involved databases, for instance the different non-anchor groups might be patients with a neutral treatment (n) of their sore throats (st) (i.e. $c_2 = (st, n)$) or stomachache problems (i.e. $c_3 = (s, n)$) or harmful treatment (w) of patients suffering from sore throats (st) (i.e. $c_4 = (st, w)$). To simplify the comparison of the databases (by providing a systematical structural guideline) the anchored database contain the (anchor) group c_1 in a specific proportion k (e.g. $k = \frac{2}{3}$) of the databases' total length. E.g. each database consisting of two thirds of successfully treated coughing patients and one third of patients with any other mutually different (symptom,outcome)-pair.

- For example a anchored database consists of 20 successfully treated coughs (i.e. c_1^{20}) and 10 neutrally treated sore throats (i.e. c_2^{10}), which results in the anchored database with 30 patients $D_1^2(\frac{2}{3}, 30) = (c_1^{20}, c_2^{10})$.
- Another database might contain 40 successfully treated coughs (i.e. c_1^{40}) and 20 neutrally treated stomachaches (i.e. c_3^{20}), i.e. anchored database $D_1^3(\frac{2}{3}, 60) = (c_1^{40}, c_2^{20})$ with 60 patients.
- Another anchored database consist of 16 successfully and 8 harmfully treated coughs (i.e. c_1^{16} and c_4^{8}), i.e. $D_1^4(\frac{2}{3}, 24) = (c_1^{16}, c_4^{8})$ with 24 patients.

The corresponding anchored chain based on $D_1^j(\frac{2}{3}, T_j)$ (for $j = 2, .., 4$ and $T_2 = 30$, $T_3 = 60$, $T_4 = 24$ and $T := \sum_{j=2}^{5} T_j = 114$) reads $F = \circ_{j=2}^{4} D_1^j(\frac{2}{3}, T_j) = (c_1^{76}, c_2^{10}, c_3^{20} c_4^{8})$.

However, within the anchor structure the comparison of the almost disjoint anchored databases $(D_i^j(k, T_j))_{j\neq i}$ is still not directly straightforward, since the precision of the anchor case c_i in each of the database varies with the corresponding lengths T_j, i.e. c_i is contained in $D_i^j(k, T_j)$ in the amount of kT_j -in our leading example reflected by the different numbers of successfully treated coughing patients. These difference in the precision would again cause the already extensively discussed difficulties in determining the average weights. In order to avoid this problem and also to respond to the issue of insufficiently precise information in non-disjoint concatenating database (see equation (3) and its derivation), we need to replicate some of the anchored database to attain a common level of precision for the anchor case. Due to the identical structure of the databases, enforcing a common precision for anchor case is equivalent to obtain a specific common length L for all involved anchored databases [15]. More precisely, for an anchored chain F of $(D_i^j(k,T_j))_{j\neq i\leq m}$ a belief induced by F should rely on an average of the beliefs induced by anchored databases $(D_i^j(k,L))_{j\neq i\leq m}$. Obviously, this enables an agent to compare easily the involved databases $D_i^j(k,L)$ since their only common case -the anchor case c_i - appears in identical amounts kL in all databases. Therefore, in comparing the anchored database (and determining the average weights) the agent can concentrate on the single and mutually different non-anchor cases.

It remains to specify and motivate a choice for a common precision level of the anchor case and (indirect) the common length L. We will introduce it in close relationship to our notion of the precision of an induced belief. As already discussed in general for non-disjoint databases in the last section (see discussion after equation (3)), there exists no replication for anchored concatenating databases, such that all single cases appear in equally precision in $(D_i^j(k,L))_{j\neq i\leq m}$ and in the related anchored chain $F = \circ_{j\neq i\leq m} D_i^j(K,T_j)$ [16]. Obviously, this leaves the freedom to choose a specific piece of information, that should be captured in equal precision in all involved induced beliefs. A very intuitive (and from our point of view most reasonable) choice to control for precision (and related confidence and cautiousness) is to ensure that the most precise and hence reliable piece of information in the anchored chain is captured in the identical precision in the beliefs induced by the corresponding replicated combining databases [17]. The focus and reliance on the most precise case can be justified by interpreting it as the driving factor of the precision of the belief. Focussing on another, less precise information would imply a less precise belief, since the most precise information would not be captured objectively anymore (in all involved databases) [18]. Hence it appears reasonable to require, that at least the most reliable information is incorporated in the belief without any distortions, which requires that it is also contained unbiased its generating (averaging) beliefs.
More technically, this can be achieved by requiring a particular adjusted length of the combining anchored databases, which is given in the following definition.

Definition 4.2
Let $F \in C^T$ be an anchored chain of $(D_i^j(k,T_j))_{j\neq i\leq m}$.
A length $L \in \mathbb{N}$ is called the ***adjusted (maximal) length*** *and denoted by $L(k,(T_j)_{j\neq i\leq m})$ if it is such that the number of observations of the most frequent case in an anchored chain $F \in C^T$ is identical to the number of observations of the most frequent case in the anchored databases $D_i^j(k,L)$ (for all $j \neq i$), (i.e. $\max_{c\in C} f_F(c)T = \max_{c\in C} f_{D_i^j(k,L)}(c)L$)* [19].

Our leading example will clarify the relationships and intuition of the adjusted length.

Example:

(i) Our doctor considers the records of different patient groups collected in two studies, i.e.

[15] This seem to be close to the EG approach in fixing the lengths of the databases. However here it is a consequence of fixing a common precision for a single case. The two approaches use different incompatible restrictions on the databases involved in the modifications of the concatenation axioms.

[16] This is due to the different appearances of the cases, i.e. for the anchored chain the appearance of an non-anchor case c_j is $(1-k)T_j$ (for all $j \neq i \leq m$) in contrast to $(1-k)L$ in the (replicated) anchored databases $D_i^j(K,T_j)$ and similar for the anchor case c_i, there exists the difference between $kT = k\sum_{j\neq i\leq m} T_j$ and kL.

[17] Another reasonable choice is the minimal precise information, that a very cautious agent might adopt (see App. E).

[18] Section 6.1 discusses another interpretation in terms of an induced persistent cautiousness attitude, that is evoked by the most precise information in the database and serves as basis for all other estimations.

[19] In Appendix E the maximum is replaced by a minimum to focus on minimal precise information.

$D_1^2(\frac{2}{3}, 30) = (c_1^{20}, c_2^{10})$ and $D_1^3(\frac{2}{3}, 60) = (c_1^{40}, c_2^{20})$ with common patient group c_1. Patient group c_1 is also the most precise information (with 60 observations) in the corresponding anchored chain $F = (c_1^{60}, c_2^{10}, c_3^{20}) \in C^{90}$. Thus the doctor requires it be matched equally precise in appropriate replications (of the study results) of the anchored databases $D_1^2(\frac{2}{3}, 30)$ and $D_1^3(\frac{2}{3}, 60)$. The adjusted length L such that for $j = 2, 3$

$$60 = \max_{c \in F} f_F(c) 90 = \max_{c \in D_1^j} f_{D_1^j}(c) L = \max\{\frac{2}{3}, \frac{1}{3}\} L = \frac{2}{3} L,$$

is given by $L = 90$, i.e. $D_1^2(\frac{2}{3}, 90) = (c_1^{60}, c_2^{30})$ and $D_1^3(\frac{2}{3}, 90) = (c_1^{60}, c_2^{30})$. Obviously, the most precise case c_1 is capture in identical precision (60) in all three databases $F, D_1^2(\frac{2}{3}, 90)$ and $D_1^3(\frac{2}{3}, 90)$. This allows an easy averaging of beliefs induced by $D_1^j(\frac{2}{3}, 90)$.

(ii) Similarly, let there be two public studies of the treatment for some specific patient groups summarized in the following anchored chain

$$F = (c_1^{30}, c_2^{40}, c_3^{80}) = (c_1^{10}, c_2^{40}) \circ (c_1^{20}, c_3^{80}) = D_1^2(\frac{1}{5}, 50) \circ D_1^3(\frac{1}{5}, 100),$$

where again the anchor patient group is "successfully treated coughs" c_1. The most precise case in F is c_3 (with 80 observations), implying that an adjusted length $L = 100$ is determined by $80 = \max\{\frac{1}{5}, \frac{4}{5}\} L$. Again, the most precise case c_3 is capture in identical precision (80) in the relevant databases F and $D_1^2(\frac{1}{5}, 100) = (c_1^{20}, c_3^{80})$, $D_1^3(\frac{1}{5}, 100) = (c_1^{20}, c_2^{80})$.

With these definitions at hand we can state our anchored Concatenation Axiom, where a modified version focussing on minimal precise information can be found in Appendix E. Recall, the length T of a database in an induced belief P becomes visible via the restriction to P_T. In particular for anchored databases $D_i^j(k, T_j)$, we can skip the length T_j in the induced belief..

Maximal Anchored Concatenation Axiom:

(i) Let $F \in C^T$ be an anchored chain of $(D_i^j(k, T_j))_{j \neq i \leq m}$, i.e. $F = \circ_{j \neq i}^m D_i^j(k, T_j)$ and let $L \in \mathbb{N}$ be the corresponding adjusted (maximal) length, i.e. $L = L(k, (T_i^j)_{j \neq i})$, then there exists $\lambda \in \Delta^m$ (where $\lambda_j = 0$ for all $j \leq m$ s. th. $T_j = 0$), such that

$$P_T(F) = \sum_{j \neq i \leq m} \lambda_j P_L(D_i^j(k))$$

(ii) Let for three distinct $i, j, l \leq m$ and any $V, W \in \mathbb{N}$: $D_i^j(1, V) = (c_i^V) \in C^V$ and $D_j^l(1/2, 2W) = (c_j^W, c_l^W) \in C^{2W}$. Let $F = D_i^j(1, V) \circ D_j^l(1/2, 2W)$, then there exist $\lambda \in int(\Delta^2)$:

$$P_{V+2W}(F) = \lambda P_{\max\{V,W\}}(D_i^j(1)) + (1 - \lambda) P_{\max\{2V,2W\}}(D_j^l(1/2)).$$

Part (i) states that the belief induced by an anchored chain is a weighted average of the beliefs induced by the related (replicated) anchored databases. The very similar and almost disjoint databases allow a simple averaging, which keeps the normative appealing spirit of the concatenation axiom. The databases share only one identical precise piece of information (the anchor case in kL-many observations)). Hence its induced identical estimate is contained in all their induced beliefs. This allows to "neglect" its impact for the determination of the average weights. Since in addition, the mutually different non-anchor cases appear only in one of the anchored databases, there emerge no difficulties in (cognitively challenging (interwoven)) balancing of differently cautious estimations based on identical, but differently precise observations in various databases. Thus, the anchored-agent can basically determine the average weights based on judging the relative importance and relevance of the mutual different non-anchor cases [20]. In this way, an anchored agent can find the average weights in a very simple case by case comparison.

The particular (maximal adjusted, Def. 4.2) length of the related corresponding concatenating databases ensures that the most precise case in an anchored chain is captured objectively in the average of their induced beliefs. An anchored-agent does not accept an average of beliefs induced by databases that evoke less precise estimations regarding this information, since this would directly imply a distortion of the precision of the belief induced by the anchored chain.

We **continue the Examples** to illustrate the anchored Concatenation axiom.
(i) cdt. The belief induced by $F = (c_1^{60}, c_2^{10}, c_3^{20})$ is an average of the beliefs induced by $D_1^2(\frac{2}{3}, 90) = (c_1^{60}, c_2^{30})$ and $D_1^3(\frac{2}{3}, 90) = (c_1^{60}, c_2^{30})$. Since by construction the estimate based on the anchor case c_1 is identically contained in all beliefs, the doctor can neglect its influence of the anchor case for determining the average weight. Hence the weights can be easily determined by just comparing the relative (a frequency-weighted) importance of c_2^{30} and c_3^{30} for evaluating the remaining parts of the anchored chain (c_2^{10}, c_3^{20}). Intuitively, the discrepancies in the precisions for c_2 and c_3 are negligible, since the focus lies predominantly on capturing perfectly the impact of the most precise case c_1. This is directly achieved in this example, since the most precise case c_1 is also the anchor case, and hence appears equally often in all databases.
(ii) cdt. The belief induced by $F = (c_1^{30}, c_2^{40}, c_3^{80})$ is an average of the beliefs induced by $D_1^2(\frac{1}{5}, 100) = (c_1^{20}, c_2^{80})$ and $D_1^3(\frac{1}{5}, 100) = (c_1^{20}, c_3^{80})$. Again, the anchor case c_1 appears equally in both replicated anchored databases, i.e. c_1^{20}, which enables to neglect it for finding the average weight. The agent only needs to weight the amount and relevance of c_2^{80} and c_3^{80} for judging (c_2^{40}, c_3^{80}). Thereby it is essential that the most precise case (in the anchored chain F) c_3 is captured perfectly. The discrepancies in the objective precisions for the cases c_1 (20 in D_1^j vs 30 in F) and c_2 (D_1^j are negligible, since the focus lies on capturing the most precise information c_3 objectively.

A straightforward consequence of the agent's focus on the most precise case and the specific structure of the anchored databases is that the estimations based on minor precise pieces of information are not made in their objective precision, but in the precision of the most precise case. This can be seen directly by the recursive application of the anchored concatenation axiom, i.e. $P_T(D) = \sum_{c \in D} \lambda_c P^c_{\max_c f_D(c) \cdot T}$ for appropriate λ_c.

[20] Of course, the estimation based on the anchor case is not contained in the same weight in each belief, but this is directly adjusted for by assigning the desired weights to the beliefs induced by the particular databases.

Since this structure (obviously) reappears in our representation theorem, we will postpone the discussion of its plausibility and reasonability to Section 5.1.

Part (ii) of the anchored Concatenation Axiom describes just a restriction to the very intuitive requirement that a belief induced by a combination of two disjoint databases should lie in between the induced beliefs of the disjoint databases separately. Averaging beliefs based on disjoint database are at the heart of the axiom, since there are no interdependencies between the information (and their precision) in the different databases. Furthermore, the axiom requires averaging only for very specific databases, i.e. a database consisting only of observations of one case and a database containing (potentially different, but) equally many observations of two other cases. The main assumption concerns the condition on the lengths, which is again driven by the agent's focus on the most precise cases, in the sense that the most precise information should be captured equally in all averaging beliefs induced by the respective databases.

4.3 Constant Similarity Axiom (for maximal anchored version)

(i) Let $F \in C^T$ be an anchored chain of $(D_i^j(k, T_j))_{j \neq i \leq m}$, i.e. $F = \circ_{j \neq i}^m D_i^j(k, T_j)$ and let $L \in \mathbb{N}$ be the corresponding adjusted (maximal) length, i.e. $L = L(k, (T_i^j)_{j \neq i})$.
If there exist some vector $\lambda \in \Delta^m$, (where $\lambda_j = 0$ for all $j \leq m$ such that $T_j = 0$) such that for some $Z \in \mathbb{N}$ the following equation holds:

$$P_{ZT}(F^Z) = \sum_{j \neq i}^{m} \lambda_j P_{LZ}(D_i^j(k)),$$

then this equation holds for all $Z \in \mathbb{N}$.
(ii) Let for three distinct $i, j, l \leq m$ and any $V, W \in \mathbb{N}$ $F = D_i^j(1, V) \circ D_j^l(1/2, 2W)$. If there exist $\lambda \in int(\Delta^2)$ for some $Z \in \mathbb{N}$ such that the following equation holds:

$$P_{Z(V+2W)}(F^Z) = \lambda P_{Z \max\{V,W\}}(D_i^j(1)) + (1 - \lambda) P_{Z \max\{2V,2W\}}(D_j^l(1/2)),$$

then this equation holds for all $Z \in \mathbb{N}$.

The average weights λs are related to (frequency weighted) relevance or similarity weights, which could in principle depend on the length of the database. However, the Constant similarity axioms allows to identify the similarity function independent of the content and the size of the databases. To require a length-independent similarity is reasonable, if the similarity values are determined by some primitive or prior knowledge about the environment, which can not be learned, influenced or based on the information contained in the database. Of course, the axiom is questionable, if an agent uses the databases not solely for evaluation of the outcome distribution, but also to learn something about structural (causal) relationship of particular features in the cases. However, the approach taken in this work excludes such deductive reasoning in deriving and updating the similarities from underlying databases [21].

[21] For deductive reasoning see also the section about the relationship to statistical methods in Section 5.4

5 Representation Theorem

5.1 Representation with maximal anchored Concatenation Axiom

Theorem 5.1

Let there be given a function $P : C^* \to \Delta(R)$. Let P_T be the restriction of P to C^T for $T \in \mathbb{N}$. Let P satisfies the Learning Axiom and the Diversity Axiom.
Then the following are equivalent:

(i) The function P satisfies the Invariance axiom, the maximal anchored Concatenation axiom, the Constant Similarity axiom
(ii) There exists for each $(T, c) \in \mathbb{N} \times C$ a unique $P_T^c \in \Delta(R)$, and a unique -up to multiplication by a positive number- function $s : C \to \mathbb{R}_+$, s. th. for all T and any $D \in C^T$:

$$P_T(D) = \frac{\sum_{c \in D} s(c) f_D(c) P^c_{T^*_D}}{\sum_{c \in D} s(c) f_D(c)} \tag{4}$$

where $T^*_D \in \mathbb{N}_+$ is defined by $T^*_D := T \cdot \max_{c \in D} f_D(c)$.

A sketch of the crucial parts of the proof can be found in Appendix B4.
The induced belief is a frequency and similarity weighted average of the estimations based on past observations. All estimations $(P^c_{\max_c f_D(c)T})_{c \in D}$ are made according to the level of cautiousness implied by the most precise case. That means, that only the most precise piece of information is captured objectively in its estimation. Hence, the axiomatized belief formation process does not achieve a perfectly objective representation (as mentioned in (2)) without any imagination effort. However such a "perfect imagination-free representation " is impossible for a sufficiently rich concatenation axiom (see Appendix E) and also carries some drawbacks (see the discussion after (2)). In any case, we are not concerned with imagining additional information to take into account objective precision.

In fact, in first place we are interested in capturing the perception of precision in form of the induced psychological effects on cautiousness and confidence. This is essential for small database containing relatively few information and is manifested in the way how estimations $P^c_{T^*_D}$ are made. From this perspective, the seemingly undesirable imagination in the axiomatized belief delivers the following intuitive and reasonable interpretation. The underlying intuition is, that an agent does not adjust constantly her cautiousness and confidence attitude in response to each differently precise information she encounters in a database. Rather, once an agent has experienced a (extreme) cautiousness and confidence feelings while estimating based on objectively available information, she keeps, adopts and transmits her developed feeling to other estimation situations. A fixed level of cautiousness according to which all estimates are made can be interpreted as an gained attitude regarding cautiousness or as a learned skill or ability to confidently estimate sufficiently cautiously. In this way, it is a sustainable reference or state of mind, which does not vanish and change for each new estimation.

For instance, an agent gained a feeling of cautiousness in the spirit of eliminating unreasonable estimates. Suppose she feels confident and considers herself cautious enough to assign only a small probability ϵ to non observed outcomes $\tilde{r} \neq r$ in estimating based on $c = (x, r)^L$. Separately, her estimation induced by $c' = (x', r')^T$ with $T > L$ assigns a slightly lower likelihood $\epsilon' < \epsilon$ to the not observed outcome $\tilde{r} \neq r'$ according to her lower cautiousness and higher confidence. Assume now, that in the past she has only estimated according to a precision level lower than L and someone tells her, that $T - L$ pieces of information c were lost and she should better estimate according to T many imagined observations. Without having experienced estimating according to higher precision T (i.e. how far she can narrow down the estimation) and being unable to imagine how she would feel if this information would be objective, she might stick to her already made estimation based on objective information c^L. However, if the agent would have estimated based on case $(c')^T$ in the past, then she has experienced her feeling of estimating according to the objective precision in $(c')^T$ and might adopt and apply the "learned" procedure how to eliminate and assign the likelihoods confidently for c^T without concerns about being too in-cautious.

The most intuitive choices for adopting a specific attitude towards cautiousness are the two extreme situations, i.e. the least and most cautious (and confident) experiences. The most precise case might come directly to her mind, because it has been observed most frequently in the database and induces an attitude of (least) cautiousness (and highest confidence) that is the basis for all estimation. In some sense the most confident and least cautious feeling outshines and distracts from any other more cautious perceptions. In contrast, the least precise information might intimidate or scare an agent and leaves a very cautious impression. She cannot be persuaded to leave her skeptical mood for a less cautious attitude that might be more appropriate for the remaining more accurate information. In our representation we focus on the optimistic view, i.e. our agent estimates according to the confidence and cautiousness gained and experienced by estimating the most precise information in the database.
In this way it is reasonable and natural to interpret the imagination of additional information in the sense of estimating according to an experienced cautiousness level or as gained skill to estimate cautiously [22].

Differences in imagined information and its imagined perception

In fact, the imagination of further additional information or more precise cases is not the cognitive difficult or challenging part in estimating based on imagined information. Think about our doctor, who just needs to imagine that the same patient enters her office again and shows the same outcome after being treated identically. Hence, the difficult part is to imagine the "correct" feeling, which would be induced by objective precision, but which is actually only existing in imagined precision. Put differently, usually the implied perception of imagined (non existing) precision differs from the perception based on objective precision. The beliefs (EG and ours) require that agents are able to ignore this difference, which might be fine if agents have experienced already a situation in which they actually estimated according to that objective precision and know her induced perception of that precision (as in our work). However, if an agent has never experienced

[22] From that perspective, our representation is even more convincing than the perfectly objective imagination-free representation (2), in which the cautiousness and confidence is altered for each case, putting the agent in different moods of cautiousness and confidence for each piece of information.

such a situation before, the requirement to imagine her feeling "correctly" (i.e. ignore the differences is cognitively challenging and psychologically confusing and can be interpreted as intentionally lying to yourself, without noticing. Does our doctor judge the treatment less cautiously after adding an imagined patient to her record?

5.2 Comparison to related belief representations

The initial motivation of EG and our paper is to modify the Concatenation axiom of BGSS to capture variations in the precision of data. A related and implied issue concerns the way how an agent is capable to deal with the problem of combining beliefs that might be based on identical, but different precise information and thus contain induced differently cautious estimates.

BGSS, EG and our work share the property that eventually the estimations involved in the final representation of a belief are subject to an unique level of precision [23]. By that, technically speaking the aggregation of different precise information is eventually not an issue. However, from an interpretational perspective, there are important differences in the motivation and reasonability of the corresponding concatenation axioms.

Consider for example the database $D = (c_1^3, c_2^4, c_3^2)$ for which a purely objective agent forms a belief according to $P(D) \in conv(\{P(c_1^3), P(c_2^4), P(c_3^2)\})$. In BGSS, the induced belief is given by $P(D) \in conv(\{P(c_1), P(c_2), P(c_3)\})$, which neglects precision and cautiousness completely. EG offers a belief $P(D) \in conv(\{P(c_1^9), P(c_2^9), P(c_3^9)\})$, where no involved estimation is made according to its objective precision. Besides the (unproblematic) imagination of additional pieces of observation for all cases, the main problematic point is the imagination on how this imagined precision is perceived, since the estimation is based on a never (not yet) experienced cautiousness level 9 (see also the discussion above). In our paper, the belief would be based on the most precise information, i.e. $P(D) \in conv(\{P(c_1^4), P(c_2^4), P(c_3^4)\})$, which also would require some (unproblematic) imagination of additional observations with respect to objective precision. However, the perception of this precision needs not to be imagined, since the agent estimates according to an already experienced precision and cautiousness level 4 (experienced for c_2).

Arad and Gayer (2012) analyze beliefs based on datasets containing imprecise pieces of information in the sense that "it is not entirely clear what occurred in them ". Roughly speaking, their approach models this sort of imprecision (ambiguity) by assuming subjective capacities. The rough relationship to the approaches discussed above is that these capacities would play the role of the probabilistic estimations occurring in the axiomatized representations of BGSS, EG and ours.

5.3 Remarks on the similarity function

One could be tempted to perceive and interpret the belief formation approaches as a translation of the question from which probability to assign to which similarity to employ. This is not completely misleading since the axiomatizations do not provide help in choosing the similarity function. This problem occurs in a similar spirit for the choice of a prior in the Bayesian approach. In the axiomatizations the similarity function is derived from

[23] in BGSS: P_∞^c for all $D \in C^*$, in EG: P_T^c for all $D \in C^T$ and here $P_{max_c f_D(c) T}$ for all $D \in C^T$.

presumably observable probability assignments given various databases. Fortunately, the similarity values need not satisfy any particular properties (even no symmetry) and hence can be derived also objectively or empirically. For example, Gilboa et al. (2006) estimate an empirical similarity function from the data by asking which similarity function best explains the observed data in a similarity-weighted frequency formula. Billot et al. (2004) axiomatized an exponential similarity function. Moreover, assigning similarities appears to be cognitively easier than stating explicit probabilities and many models in the psychology and computer-science literature deal with determination of similarity measures (e.g. Tversky (1977), Schank (1986), Heit, Heit and Rubinstein (1994), Goldstone and Son (2005)).

5.4 Remarks on relationship to statistical methods

In the introduction we mentioned already the relationship between the axiomatic approaches to belief formation in the data-based information structure and statistical approaches like inferences. In this section we want to discuss shortly similarities and differences to existing statistical methods. Obviously, the versions of the concatenation axioms and the derived representations satisfies the following special cases of frequentism. For $s(x_i, x_t) = 1$, our belief formation coincides with the simple average or frequentist approach, if we identify with P^c a Dirac measure on the actually observed outcome. However, the conditional frequentist cannot be covered since the corresponding $s(x_i, x_t) = 1_{\{x_t = x_i\}}$ is not strictly positive (as required), but Bleile (2014) offers a modification that captures it. Gilboa et al (2010, 2011) and EG show the compatibility with other statistical methods, like kernel estimation and classification (e.g. assign x to either class a or b: define $s(a, (x_c, a_c)) = k(x_c, x)1_{\{a=a_c\}}$ using a kernel function k). As discussed in more detail in Gilboa et al. (2010) p. 16f, the framework can be also employed in contexts, where the observations (e.g. cases) and the prediction (e.g. possible theories) are structurally disjoint. For instance ranking theories by log likelihood methods $s(t, c) = log(p(c|t))$ is also possible where t represent a theory and $p(c|t)$ denotes the conditional likelihood of case c if theory t is true.

However, the main difference to statistical inference is that the axiomatic approaches are concerned with inductive reasoning and do not allow for deductive reasoning, which is the issue of traditional statistical regression approaches. Let there be a database consisting of observation $D = ((x_i, r_i)_{i \leq n})$ and a new problem x_t. A regression approach would try to learn the (empirical) similarity weights $(s(x_i, x_t))_i$ that best explains the database by best fitting an estimate of r_j for all $j \leq n$ and $r_j^s = \frac{\sum_{i \neq j} s(x_i, x_j) r_i}{\sum_i s(x_i, x_j)}$ (see also Gilboa et al. (2006)). Hence in a statistical regression context the weights s are deduced endogenously via the observed data and are updated with new observation, i.e. the weights would be database dependent. Put differently, linear regression analysis (and empirical similarities) use deductive reasoning to derive the weights and then apply them inductively to infer the prediction. In contrast, the constant similarity (and the concatenation) axiom requires that the weights are fixed and database independent, i.e. there is no updating or learning of the weights.

However, the axiomatization of a belief formation (in close relationship to statistical methods) is still meaning- and insightful, since it allows to inspect, how plausible, consistent and sensible (in the sense of normative appealing axioms) asymptotic statistical methods are also for small database and its implied precision related concerns. From this perspective, axiomatizations suitable for small databases (as done here) play an important role in order to find a sound foundation of statistical methods in non-asymptotic contexts.

6 Conclusion

The paper deals with the question how agents form beliefs explicitly in an environment with limited, heterogenous and differently precise information that cannot be condensed into a widely used (perfect) state space a la Savage. We axiomatize a belief formation that can be interpreted as a generalized subjective frequentist approach that incorporates subjective perceptions regarding the relevance and precision of the information in the database. We identify increasing precision of information by additionally observed pieces of confirming information.

Our work is based on the axiomatization of a belief in BGSS that neglects the potential impacts of differently precise information. Thereby, their belief formation is most suitable for sufficiently large databases and less reasonable for small databases, which are captured by our approach. Their belief formation implies that an agent is able to perfectly learn from observations in a very objective and instantaneously way, without displaying any sense of cautiousness and concerns about being potentially mistaken. Our axiomatized cautious belief focusses on precision related cautiousness and confidence in the predictions. The different versions of the main concatenation axiom in the approaches of BGSS, EG and ours describe the relationships between databases and their induced beliefs.

In the context of caring for precisions in a cautious belief formation an agent following the concatenation axiom of BGSS and EG's version would be faced by immense cognitive problems to handle and compare differently precise pieces of information contained in different databases. Our modification and restriction of the axiom takes into account these precision related cognitive problems in describing the relationships. This is achieved by requiring that agents only need to be capable to determine the relationship between databases and their induced beliefs for specifically structured (almost disjoint) databases that allow an cognitively easy comparison (without precision and cautiousness concerns). Moreover, it states that an agent controls for precision and its perceptional impacts in a cautious belief by capturing the most precise (and hence reliable) information objectively in its induced belief.

The resulting cautious belief is a weighted sum of cautious estimates induced by past observed information. The weights are determined by frequencies of the observed cases and their similarities with the problem under consideration. The induced estimates depend on a cautiousness level implied by the most precise case, which can be interpreted as the appropriate (gained) attitude regarding cautiousness in this database.

A Proof of Theorem 6.1: Necessity part, i.e. (ii) $\Rightarrow$ (i)

We need to show that the representation (4) satisfies the axioms, where the Invariance axiom is obviously met.

For the **Maximal Anchored Concatenation axiom, part (i)**,

let $D \in C^T$ be a chain of $D_i^j(k, T_j) = (c_j^{(1-k)T_j}, c_i^{kT_j})$ for all $j \neq i \leq |C|$ and $T := \sum_{j \neq i} T_j$, i.e.
$D = \circ_{j \neq i} D_i = (c_1^{(1-k)T_1}, c_2^{(1-k)T_2}, ..., c_{i-1}^{(1-k)T_{i-1}}, c_i^{kT}, c_{i+1}^{(1-k)T_{i+1}}, ..., c_{|C|}^{(1-k)T_{|C|}})$.
Let $L = L(k, (T_j)_{j \neq i})$ be the corresponding adjusted length. Hence, we have
$f_D = (\frac{(1-k)T_1}{T}, \frac{(1-k)T_2}{T}, ..., \frac{(1-k)T_{i-1}}{T}, k, \frac{(1-k)T_{i+1}}{T}, ..., \frac{(1-k)T_{|C|}}{T})^t$ and
$f_{D_i^j(k,ZT_j)} = (0, ..., 0, (1-k), 0, .., 0, k, 0, ..., 0)^t$.

Observe that $f_{D_i^j(k,T_j)} = f_{D_i^j(k,ZT_j)}$ and hence we will abbreviate $f_{D_i^j(k,ZT_j)}$ by $f_{D_i^j(k)}$.
We get:

$$\begin{aligned}
P_T(D) &= \frac{\sum_{c \in C} s(c) f_D(c) P^c_{\max_{c \in D} f_D(c) \cdot T}}{\sum_{c \in C} s(c) f_D(c)} \\
&= \frac{1}{\sum_{c \in C} s(c) f_D(c)} \cdot \Big(\sum_{j \neq i} s(c_j) \frac{(1-k)T_j}{T} P^{c_j}_{\max_{c \in D} f_D(c) \cdot T} + s(c_i) \frac{\sum_{j \neq i} kT_j}{T} P^{c_i}_{\max_{c \in C} f_D(c) \cdot T} \Big) \\
&= \frac{1}{\sum_{c \in C} s(c) f_D(c)} \cdot \Big(\sum_{j \neq i} \Big[s(c_j) \frac{(1-k)T_j}{T} P^{c_j}_{\max_{c \in C} f_D(c) \cdot T} + s(c_i) \frac{kT_j}{T} P^{c_i}_{\max_{c \in C} f_D(c) \cdot T} \Big] \Big) \\
&= \frac{1}{\sum_{c \in C} s(c) f_D(c)} \cdot \Big(\sum_{j \neq i} \frac{T_j}{T} \sum_{c \in C} s(c) f_{D_i^j(k)}(c) P^c_{\max_{c \in C} f_D(c) \cdot T} \Big[\frac{\sum_{c \in C} s(c) f_{D_i^j(k)}(c)}{\sum_{c \in C} s(c) f_{D_i^j(k)}(c)} \Big] \Big)
\end{aligned}$$

To proceed, we need to specify $\max_{c \in C} f_D(c) \cdot T$, which is by definition of the adjusted length L exactly equal to $\max_{c \in C} f_{D_i^j}(c) L$, hence:

$$\begin{aligned}
P_T(D) &= \frac{1}{\sum_{c \in C} s(c) f_D(c)} \cdot \Big(\sum_{j \neq i} \frac{T_j}{T} \sum_{c \in C} s(c) f_{D_i^j(k)}(c) \frac{\sum_{c \in C} s(c) f_{D_i^j(k)}(c) P^c_{\max_{c \in C} f_{D_i^j}(c) L}}{\sum_{c \in C} s(c) f_{D_i^j(k)}} \Big) \\
&\overset{(*)}{=} \frac{1}{\sum_{j \neq i} \frac{T_j}{T} \sum_{c \in C} s(c) f_{D_i^j(k)}(c)} \cdot \Big(\sum_{j \neq i} \frac{T_j}{T} \sum_{c \in C} s(c) f_{D_i^j(k)}(c) P_L(D_i^j(k, L)) \Big) \\
&= \sum_{j \neq i} \lambda_j P_L(D_i^j(k, L))
\end{aligned}$$

where we used $\sum_{c \in C} s(c) f_D(c) = \sum_{j \neq i} \frac{T_j}{T} \sum_{c \in C} s(c) f_{D_i^j(k)}(c)$ in $(*)$.
From the last equation we get for all $j \neq i \leq |C|$

$$\lambda_j = \frac{\frac{T_j}{T} \sum_{c \in C} s(c) f_{D_i^j(k)}(c)}{\sum_{j \neq i} \frac{T_j}{T} \sum_{c \in C} s(c) f_{D_i^j(k)}(c)} \tag{5}$$

Hence the first part of the anchored concatenation axiom is satisfied.

For the **Maximal Anchored Concatenation axiom, part (ii)**:
let w.l.o.g. $D_i^j(1, T) = D_1^2(1, T) = (c_1^T) \in C^T$ and $D_j^l(1/2, 2W) = D_2^3(1/2, 2W) =$

$(c_2^W, c_3^W) \in C^{2W}$, then we need to show:

$$P_{T+2W}(D_1^2(1,T) \circ D_2^3(1/2,2W)) = \lambda P_{\max\{T,W\}}(D_1^2(1,\max\{T,W\})) \\ + (1-\lambda)P_{\max\{2T,2W\}}(D_2^3(1/2, 2\max\{T,L\}))$$

We have $P_{\max\{T,W\}}(D_1^2(1,\max\{T,W\}) = P^{c_1}_{\max\{T,W\}}$
Since P satisfies the maximal anchored concatenation axiom part (i) we have for $D_2^3(1/2,2\max\{T,W\}) = D_1^2(0,\max\{T,W\}) \circ D_1^3(0,\max\{T,W\}) = (c_2)^{\max\{T,W\}} \circ (c_3)^{\max\{T,W\}}$ with the adjusted length L such that $\frac{1}{2}2\max\{T,W\} = L$, i.e. $L(0,\max\{T,W\},\max\{T,W\}) = \max\{T,W\}$, that there exist some $\lambda \in (0,1)$ such that

$$P_{\max\{2T,2W\}}(D_2^3(1/2,2\max\{T,W\})) = \lambda P^{c_2}_{\max\{T,W\}} + (1-\lambda)P^{c_3}_{\max\{T,W\}}$$

Hence, using this, we get for the maximal anchored concatenation axiom (ii) the following representation for some $\lambda \in \Delta^3$:

$$P_{T+2W}(D_1^2(1,T) \circ D_2^3(1/2,2W)) = \sum_{i=1}^{3} \lambda_i P^{c_i}_{\max\{T,W\}}$$

But this is obviously satisfied by the representation (4) in the Theorem 5.1, since for the frequency vector $f_{D_1^2(1,T)\circ D_2^3(1/2,2W)} = (\frac{T}{T+2W}, \frac{W}{T+2W}, \frac{W}{T+2W}, 0, ..., 0)^t$, we have $\max_{c\in C} f_{D_1^2(1,T)\circ D_2^3(1/2,2L)}(c)(T+2W) = \max\{T,W\}$, and hence $P_{T+2W}(D_1^2(1,T) \circ D_2^3(1/2,2W)) = \frac{\sum_{i=1}^3 s(c_i) f_{D_1^2(1,T)\circ D_2^3(1/2,2W)}(i) P^{c_i}_{\max\{T,W\}}}{\sum_{i=1}^3 s(c_i) f_{D_1^2(1,T)\circ D_2^3(1/2,2W)}(c_i)}$, i.e. for $i = 1,2,3$

$$\lambda_i = \frac{s(c_1) f_{D_1^2(1,T)\circ D_2^3(1/2,2W)}(c_i) P^{c_1}_{\max\{T,W\}}}{\sum_{i=1}^3 s(c_i) f_{D_1^2(1,T)\circ D_2^3(1/2,2W)}(c_i)} \tag{6}$$

Hence the (ii)-part of the maximal anchored concatenation axiom is also satisfied.

The **Constant similarity axiom** is also satisfied, which can be shown by adopting the above proof for the concatenation axiom.

Replacing $D = \circ_{j\neq i}^{|C|} D_i^j(k,T_j)$, where $\sum_{j\neq i}^{|C|} T_j = T$ by $D^Z = \circ_{j\neq i}^{|C|} D_i^j(k, ZT_j)$ in the proof of the maximal anchored concatenation axiom part (i) and transform equation (5), we get

$$\lambda_j^Z = \frac{\frac{ZT_j}{ZT} \sum_{c\in C} s(c) f_{D_i^j(k,ZT_j)}(c)}{\sum_{j\neq i} \frac{ZT_j}{ZT} \sum_{c\in C} s(c) f_{D_i^j(k,ZT)}(c)} = \lambda_j,$$

where in the last equation $f_{D_i^j(k,ZT)}(c) = f_{D_i^j(k)}(c)$ is used.
For part (ii), analogous reasoning using equation (6) yields the desired result.

Therefore the similarity axiom is satisfied, which completes the proof of the Theorem 5.1 direction (ii) implies (i).

□

Now we will focus on the the direction (i) implies (ii).

B Proof of Theorem 1: Sufficiency part, i.e. (i) ⇒ (ii)

The proof will exploit the fact that by the invariance axiom it is possible to rewrite the database framework into a frequency framework. This allows to work on simplex instead with lists of cases or databases. In the following, we need to translate the database structure to a frequency terminology.

B.1 Translation into frequency framework

By the Invariance axiom each database $D \in \mathcal{C}^T$ can be identified by a pair (f_D, T), where $f_D \in \Delta(C)$ represents a frequency vector of appearances of cases in the database D and T is the length of the database.
Note, that in the frequency setup, without knowing the corresponding database D to which a frequency vector $f \in \Delta(C)$ (should) belong (in the sense of representing this specific database D), the frequency vector can be linked to infinitely many databases D^Z for all $Z \in \mathbb{N}_+$. Hence one needs to link frequency and the length of the database.

The following set represents all frequency vectors corresponding to databases $D \in C^T$:

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

Observe that if $f \in \Delta_T(C)$, then $f \in \Delta_{TZ}(C)$ for all $Z \in \mathbb{N}_+$.
Since the set of cases C is fixed, we reduce the notational effort and will abbreviate $\Delta_T(C)$ by Δ_T, i.e. Δ_T denotes the set of all frequency vectors representing databases of length T and the set of all rational frequency vectors on C is denoted by Δ.
Hence by the **Invariance axiom** each $D \in C^T$ can be represented by a $f \in \Delta_T$, where again $f(i) := f_D(c_i)$ denotes the frequency of case c_i for all $i \leq |C|$.

Definition of the belief on frequencies:
From now on we consider only probabilities $P \in \Delta(R)$ that satisfy the invariance axiom, i.e. then the definition of P on databases translate to P defined on frequency vectors in the following way:
For all $f \in \Delta$ define the function P and its restriction to P_T for all $T \in \mathbb{N}$ on frequency vectors by
$P : X \times \Delta \;\to\; \Delta(R)$ such that $P(f) := P(D)$ for $f \in \Delta$ and $D \in C$ related by $f = f_D$.

$P_T :\; X \times \Delta_T \;\to\; \Delta(R)$ such that $P_T(f) := P_T(D)$ for $f \in \Delta_T$ and $D \in C^T$ related by $f = f_D$.
As long as no length is fixed, $f \in \Delta$ is universal and the length T of the database, which f represents becomes visible only in the restriction of $P(f)$ to the specific $P_T(f)$, i.e. P_T pins down the unique database the frequency vector is able to represent, namely the database with length T. Of course, under the condition that the frequency vector allows the existence of such a database in this specific length.
Recall, we assume $|C| = m$.

Notation:

(i) For all $j \in \{1, 2, ..., m\}$ denote by f^j the j-th unit vector in $\mathbb{R}^m$, i.e. the frequency vector representing a database containing only cases $c_j \in C$, hence an extremal point in Δ, i.e. $f^j = (0, ..., 0, \underbrace{1}_{j-th}, 0, ..., 0)^t$

(ii) The frequency vector corresponding to the anchored database $D_i^j(k, T) = (c_j^{(1-k)T}, c_i^{kT})$ is given by

$$f_{D_i^j(k,T)} = (0, ...0, \underbrace{(1-k)}_{j-th}, 0, .., 0, \underbrace{k}_{i-th}, 0, ..., 0)^t$$

Since $f_{D_i^j(k,W)} = f_{D_i^j(k,T)}$ for all T and W, the length is totally immaterial for the frequency vector and hence neglected from now on, i.e. the frequency vector corresponding to the anchored databases $D_i^j(k,T)$ for all $j \neq i \leq m$ is denoted for all T such that $kT \in \mathbb{N}$ by

$$f_i^j(k) := f_{D_i^j(k,T)}$$

Note that $f_i^j(k)$ is still the whole frequency vector, i.e. $f_i^j(k) \in \Delta$, whereas $f_i^j(k)(l)$ represents the l-th component of the vector and refers to the frequency of case c_l, i.e. $f_i^j(k)(l) \in [0,1] \cap \mathbb{Q}$.

Definition

The frequency vector corresponding to a database $D \circ D^*$ with corresponding frequency vectors $f_D = f \in \Delta_T$ and $f_{D^*} = f^* \in \Delta_W$ is denoted by $f \circ f^*$ and given by

$$f \circ f^* = \Big(\frac{f(1)T + f^*(1)W}{T+W}, ..., \frac{f(m)T + f^*(m)W}{T+W}\Big)^t \in \Delta_{T+W}$$

Now we need to translate the axioms into this frequency framework.

B.2 Axioms in the frequency framework

Maximal anchored Concatenation Axiom:
(i) Let there be $f \in \Delta_T$, for all $j \neq i \leq m$ $f_i^j(f(i)) \in \Delta_{T_j}$ and $\sum_{j\neq i}^m T_j = T$ such that for $(\alpha_j)_{j\neq i \leq m} \in [0,1]$ and $\sum_{j\neq i}^m \alpha_j = 1$, i.e. $f = \sum_{j\neq i}^m \alpha_j f_i^j(f(i))$.
Let $L = L(f(i), (T_j)_{j\neq i\leq m}) \in \mathbb{N}$ be the corresponding adjusted length, i.e. $\max_{i\leq m} f(i)T = \max_{l\leq m} f_i^j(f(i)))l)L$.
Then there exist $\lambda \in \Delta^{m-1}$ (where $\lambda_j = 0$ for all $j \neq i \leq m$ such that $\alpha_j = 0$), such that

$$P_T(f) = \sum_{j\neq i\leq m} \lambda_j P_L(f_i^j(f(i))$$

(ii) Let for distinct $i,j,l \leq m$ $f_i^j(1) \in \Delta_T$ and $f_j^l(1/2) = (0,...0,1/2,0,...,0,1/2,0,..,0)^t \in \Delta_{2W}$, then there exists a $\lambda \in \Delta^2$, such that:

$$P_{T+2W}(f_i^j(1) \circ f_j^l(1/2)) = \lambda P_{\max\{T,W\}}(f_i^j(1)) + (1-\lambda) P_{\max\{2T,2W\}}(f_j^l(1/2))$$

Constant Similarity Axiom:
(i) Let there be $f \in \Delta_T$, for all $j \neq i \leq m$ $f_i^j(f(i)) \in \Delta_{T_j}$ such that for $(\alpha_j)_{j\neq i\leq m} \in [0,1]$ and $\sum_{j\neq i}^m \alpha_j = 1$, i.e. $f = \sum_{j\neq i}^m \alpha_j f_i^j(f(i))$.
Let $L = L(f(i), (T_j)_{j\neq i\leq m}) \in \mathbb{N}$ be the corresponding adjusted length.
If there exist $\lambda \in \Delta^{m-1}$ (where $\lambda_j = 0$ for all $j \neq i \leq m$ such that $\alpha_j = 0$), such that for some $Z \in \mathbb{N}_+$

$$P_{ZT}(f) = \sum_{j\neq i\leq m} \lambda_j P_{ZL}(f_i^j(k)),$$

then the equation holds for all $Z \in \mathbb{N}_+$.

(ii) Let for distinct $i,j,l \leq m$ $f_i^j(1) \in \Delta_T$ and $f_j^l(1/2) = (0,...0,1/2,0,...,0,1/2,0,..,0)^t \in \Delta_{2W}$, then there exists a $\lambda \in \Delta^2$, such that for for some $Z \in \mathbb{N}_+$

$$P_{Z(T+2W)}(f_i^j(1) \circ f_j^l(1/2)) = \lambda P_{Z\max\{T,W\}}(f_i^j(1)) + (1-\lambda) P_{Z\max\{2T,2W\}}(f_j^l(1/2)),$$

then the equation holds for all $Z \in \mathbb{N}_+$.

Learning Axiom:
For all $i \in \{1, 2, ..., C\}$: $(P_T(f^i))_{T \in \mathbb{N}_+}$ converges to $P_\infty(f^i) = P^i_\infty$.

Diversity Axiom:
There exist some $T^* \in \mathbb{N}_+$, such that for all $T \geq T^*$, no three elements of $\{(P_T(f^j))_{j \leq m}\}$ are collinear.

Before stating the unproofed direction of Theorem 5.1 in the frequency version, we will present some helpful remarks and lemmas.

B.3 Useful Observations

Remark B.1
Explicitly the adjusted length defined in Definition 4.2 is given by:

$$L(k, T_1, T_2, ..., T_m) = \begin{cases} \max_j\{T_j\} & \textit{if } k \leq \frac{\max_j\{T_j\}}{\max_j\{T_j\}+T} =: k^* \in (\frac{1}{m+1}, \frac{1}{2}) \\ \frac{k}{1-k}T & \textit{if } k \in (k^* = \frac{\max_j\{T_j\}}{\max_j\{T_j\}+T}, \frac{1}{2}) \\ T & \textit{if } k \geq \frac{1}{2} \end{cases}$$

Remark B.2
For all $f \in \Delta$ there exist for all anchor case $c_i \in C$, for $i \leq m$, a decomposition $f = \sum_{j \neq i} \alpha^j_i f^j_i(f(i))$, where for all $j \neq i \leq m$, $\alpha^j_i \in [0, 1]$ are given by $f(j) = \alpha^j_i(1 - f(i))$.

Note that α^j_i corresponds to the (relative) sizes of the databases $D^j_i(f(i), \cdot)$ (corresponding to the particular frequency vectors $f^j_i(f(i))$) in relation to the specific database D (which is represented by the frequency vector f). For instance, assume that $f^j_i(f(i)) \in \Delta_{V_j}$ and $\sum_{j \neq i} V_j = V$, then $f \in \Delta_V$ and $\alpha^j_i = \frac{V_j}{V}$.

In general for all $j \neq i \leq m$, $f^j_i(f(i))$ can represent a database with length $t \cdot \widetilde{T^j_i}$, where $t \in \mathbb{N}$ and $\widetilde{T^j_i} \in \mathbb{N}$ is the smallest length W such that $f(i)W$ is a natural number (and hence also $(1 - f(i))W \in \mathbb{N}$).

To specify the (smallest) length $Z_i \in \mathbb{N}$ of the database D corresponding to the decomposition of f via anchor case c_i, i.e. $f = \sum_{j \neq i} \alpha^j_i f^j_i(f(i))$, we extend all $\widetilde{T^j_i} \in \mathbb{N}$ with the smallest $z^j_i \in \mathbb{N}$ such that for $j \neq i \leq m$ all α^j_is are the fractions with the smallest common denominator Z_i, i.e. $\alpha^j_i = \frac{z^j_i \widetilde{T^j_i}}{Z_i}$. In this way the smallest lengths of the databases represented by $f^j_i(f(i))$, that can be used for the decomposition of f via anchor c_i, are exactly given by

$$T^j_i := z^j_i \widetilde{T^j_i} = \alpha^j_i Z_i \in \mathbb{N} \quad \text{and} \quad Z_i = \sum_{j \neq i \leq m} T^j_i \tag{7}$$

Hence $f \in \Delta_{Z_i}$ and for all $j \neq i \leq m$ the $f^j_i(f(i)) \in \Delta_{T^j_i}$.
Obviously, choosing a different anchor case $c_l \in C$ for the decomposition of f will lead to a different smallest denominator Z_l (and induced length of database which f represents) and different lengths of the databases $D^j_l(f(l), T^j_l)$ for all $j \neq l \leq m$, which are represented by $f^j_l(f(l)) \in \Delta_{T^j_l}$.

Definition B.1
For all $f \in \Delta$ and $i \leq m$, the decomposition $f = \sum_{j \neq i} \alpha^j_i f^j_i(f(i))$, where $f^j_i(f(i)) \in \Delta_{T^j_i}$

and T_i^j are as in (7), is called the smallest decomposition of f via anchor case $c_i \in C$, which represents a database of length $Z_i = \sum_{j \neq i \leq m} T_i^j$ and is denoted by $(f_i^j(f(i)), T_i^j)_{j \neq i}$.

The following Lemma shows consistency of the axiomatization with respect to the possible smallest decompositions based on different anchor cases.

Lemma B.1
Let $m \geq 3$. If $P : X \times \Delta \to \Delta(R)$ and its restriction P_T to $X \times \Delta$ satisfies the maximal anchored concatenation axiom and the constant similarity axiom. Then, P is well defined or consistent with respect to the different possible smallest decomposition and for all $T \geq 2$ and any $f \in \Delta_T$

$$P_T(f) = \sum_{j \leq m} \lambda_j P_{max_{i \leq m} f(i) T}(f^j) \tag{8}$$

Proof:

For all $f \in \Delta$, there exists the smallest decomposition via anchor case $c_i \in C$ (as in Definition B.1) $(f_i^l(f(i)), T_i^l)_{l \neq i}$, where $f(l) = \frac{T_i^l}{Z_i}(1 - f(i))$ (which implies $(1 - f(i))T_i^l = f(l)Z_i$). We have to show that independent of the choice of the anchor case $c_i \in C$, the induced belief $P_T(f)$ is identical for all $T \in \mathbb{N}$ such that $f \in \Delta_T$.
We differentiate into the three situations of adjusted lengths given in Remark B.1, which are based on the different frequencies of chosen anchor case $c \in C$.

(i) Let $f(i) \leq k^*$, assume w.l.o.g. that $\max_{l \leq m} f(l) = f(j)$, hence $\max_{l \leq m} T_i^l = T_i^j$:
Applying the maximal anchored concatenation axiom in a first step for $k = f(i) \leq \frac{\max_{j \neq i} T_i^j}{\max_{j \neq i} T_i^j + Z_i}$ with adjusted length $L(f(i), (T_i^l)_{l \neq i}) = T_i^j$, and in the second line for $k = 0$ with adjusted length $L(0, f(i)T_i^j, (1 - f(i))T_i^j) = (1 - f(i))T_i^j$ and for some $a \leq m$ such that $a \neq i, l$, we get for some $\lambda, \gamma, \beta \in \Delta^m$:

$$\begin{aligned}
P_{Z_i}(f) &= \sum_{l \neq i} \lambda_l P_{L(f(i),(T_i^l)_{l \neq i})}(f_i^l(f(i))) = \sum_{l \neq i} \lambda_l P_{T_i^j}(f_i^l(f(i))) \\
&= \sum_{l \neq i} \lambda_l (\gamma_l P_{L(0,f(i)T_i^j,(1-f(i))T_i^j)}(f_a^i(0)) + (1 - \gamma_l) P_{L(0,f(i)T_i^j,(1-f(i))T_i^j)}(f_a^l(0))) \\
&= \sum_{l \neq i} \lambda_l (\gamma_l P_{(1-f(i))T_i^j}(f^i) + (1 - \gamma_l) P_{(1-f(i))T_i^j}(f^l)) \\
&= \sum_{l} \beta_l P_{(1-f(i))T_i^j}(f^l) = \sum_{l} \beta_l P_{\max_{l \leq m} f(l) Z_i}(f^l)
\end{aligned}$$

Now by the constant similarity axiom, we get also that $P_T(f) = \sum_l \beta_l P_{\max_{l \leq m} f(l) T}(f^l)$ for all T such that $f \in \Delta_T$.

(ii) $f(i) \in (k^*, 1/2)$, which implies that $f(i) \geq (1 - f(i)) \max_{j \neq i} T_j / T = \max_{j \neq i} j f(j)$, i.e. $max_{l \leq m} f(l) = f(i)$:
With $L(f(i), (T_j)_{j \neq i}) = \frac{f(i)}{1-f(i)} Z_i$:

$$\begin{aligned}
P_{Z_i}(f) &= \sum_{l \neq i} \lambda_l P_{L(f(i),(T_l)_{l \neq i})}(f_i^l(f(i))) = \sum_{l \neq i} \lambda_l P_{\frac{f(i)}{1-f(i)} Z_i}(f_i^l(f(i))) \\
&= \sum_{l \neq i} \lambda_l (\gamma_l P_{L(0,f(i)\frac{f(i)}{1-f(i)} Z_i,(1-f(i))\frac{f(i)}{1-f(i)} Z_i)}(f^i) + (1 - \gamma_l) P_{L(0,f(i)\frac{f(i)}{1-f(i)} Z_i,(1-f(i))\frac{f(i)}{1-f(i)} Z_i)}(f^l)) \\
&= \sum_{l \neq i} \lambda_l (\gamma_l P_{f(i) Z_i}(f^i) + (1 - \gamma_l) P_{f(i) Z_i})(f^l) = \sum_{l} \beta_l P_{\max_{l \leq m} f(l) Z_i}(f^l)
\end{aligned}$$

Again by the constant similarity axiom, we get that $P_T(f) = \sum_l \beta_l P_{\max_{l\le m} f(l)T}(f^l)$ for all T such that $f \in \Delta_T$.

(iii) $f(i) \geq 1/2$, i.e. $f(i)$ is maximal frequency, which gives $L(f(i),(T_l)_{l\neq i}) = Z_i$, hence

$$\begin{aligned} P_{Z_i}(f) &= \sum_{l\neq i} \lambda_l P_{L(f(i),(T_l)_{l\neq i})}(f_i^l(f(i))) = \sum_{l\neq i} \lambda_l P_{Z_i}(f_i^l(f(i))) \\ &= \sum_{l\neq i} \lambda_l(\gamma_l P_{L(0,f(i)Z_i,(1-f(i))Z_i)}(f^i) + (1-\gamma_l) P_{L(0,f(i)Z_i,(1-f(i))Z_i)}(f^l)) \\ &= \sum_{l\neq i} \lambda_l(\gamma_l P_{f(i)Z_i}(f^i) + (1-\gamma_l) P_{f(i)Z_i}(f^l)) = \sum_l \beta_l P_{\max_{l\le m} f(l)Z_i}(f^l) \end{aligned}$$

Again by the constant similarity axiom, we get that $P_T(f) = \sum_l \beta_l P_{\max_{l\le m} f(l)T}(f^l)$ for all T such that $f \in \Delta_T$. □

Remark B.3
In the proof above, of course, it would be sufficient to prove (i) and combine it with the fact that for all $f \in \Delta$, there exist $i \neq j \leq m$, such that $f(i) \leq k^*$, otherwise for $l \leq m$ $f(l) \geq \frac{\max_{j\neq i} T_i^j}{\max_{j\neq i} T_i^j + Z_i}$ and hence $\sum_{l\le m} f(l) \geq (n+1)\frac{\max_{j\neq i} T_i^j}{\max_{j\neq i} T_i^j + Z_i} \geq 1$ since $Z_i \leq n(\max_{j\neq i} T_i^j)$. But it would not show directly the consistency wrt. the different particular decompositions.

Remark B.4
Let $f \in \Delta$ be expressed as convex combination of the set $\{f_1, f_2, f_3\}$ for some $f_i \in \Delta_T$ for all $i = 1,2,3$, i.e. $f = \beta_1 f_1 + \beta_2 f_2 + (1-\beta_1-\beta_2) f_3$.
As in Remark B.2 we apply the relativ length interpretation of the weights $\beta_i \in (0,1)$ for all $i = 1,2,3$, to get the (potentially) smallest induced length H of the database represented by f via the convex combination of databases $D_i \in C^T$, which are represented by $f_i \in \Delta_T$. That is, H is again the smallest possible denominator of all β_i such that for all $i = 1,2,3$ $\beta_i = \frac{z_i T}{H}$ for some $z_i \in \mathbb{N}$ and hence we have that $f \in \Delta_H$ can be combined by the decomposition $(f_i)_{i\le 3}$, where $f_i \in \Delta_{\beta_i H = z_i T}$ for $i = 1,2,3$.

The following Lemma mirrors Lemma A.4 in EG.

Lemma B.2
Let P satisfy the maximal anchored concatenation and constant similarity axiom. For $m \geq 3$ let $(s_j)_{j\le m}$ be a collection of similarity weights. Define the function $P^s : X \times \Delta(C) \to \Delta(R)$ and for any $T \in \mathbb{N}$, $T \geq 2$ and any $f \in \Delta_T$ the restriction P_T^s to $X \times \Delta_T$ by

$$P_T^s(f) = \frac{\sum_{j\le m} s_j f(j) P_{\max_{j\le m} f(j)T}(f^j)}{\sum_{j\le m} s_j f(j)}$$

Suppose that for some $T \geq T^*$ and $f \in \Delta_T$ it holds $P_T(f) = P_T^s(f)$.
Then, $P_W(f) = P_W^s(f)$ for all $W \in \mathbb{Z}$ such that $f \in \Delta_W$

Proof:

Let $T(f)$ be the smallest T such that $f \in \Delta_{T(f)}$, this implies that for all $l \in \mathbb{N}$ $f \in \Delta_{lT(f)}$.
By Lemma B.1 we know that P can be represented as in representation (8) , hence we get the following.
If there exist some $\lambda \in \Delta^m$ (with $\lambda_i = 0$ if and only if $f(i) = 0$) such that it satisfies for some $l \in \mathbb{N}_+$,

$$P_{lT(f)}(f) = \sum_{j=1}^{m} \lambda_j P_{max_{i\le m} f(i)T(f)}(f^j),$$

then by the constant similarity axiom it also holds for all $l \in \mathbb{N}_+$. In particular, for l such that $lT(f) = T$ the following holds:

$$P_T(f) = \sum_{j=1}^{m} \lambda_j P_{max_{i\leq m} f(i)T(f)}(f^j) \overset{by\,ass}{=} \frac{\sum_{j=1}^{m} s_j f(j) P_{max_{i\leq m} f(i)T}(f^j)}{\sum_{j=1}^{m} s_j f(j)} = P_T^s(f)$$

By the Diversity axiom's non-collinearity condition, we get $\lambda_j = \frac{s_j f(j)}{\sum_{j=1}^{m} s_j f(j)}$.

Since $P_{lT(f)}^s = \frac{\sum_{j=1}^{m} s_j f(j) P_{max_{i\leq m} f(i)lT(f)}(f^j)}{\sum_{j=1}^{m} s_j f(j)} = \sum_{j=1}^{m} \lambda_j P_{max_{i\leq m} f(i)T(f)}(f^j) = P_{lT(f)}(f)$ for all l, the proof is completed. $\square$.

B.4 Theorem 5.1 (i) ⇒ (ii) in frequency version

Theorem B.1

Let there be given a function $P : X \times \Delta \to \Delta(R)$. Let P_T the restriction of P to $X \times \Delta_T$ and let for $T \geq 2$ $P_T : \Delta_T \to \Delta(R)$ satisfy the following conditions

(i) Learning Axiom
(ii) Diversity Axiom
(iii) Maximal Anchored Concatenation Axiom
(iv) Constant Similarity Axiom

Then, for all $T \geq 2$, there exist unique probability vectors $(P_T^j)_{j\leq C} \in \Delta(R)$, and unique -up to multiplication by a strictly positive number- positive numbers $(s_j)_{j\leq m} \in \mathbb{R}^+$, such that for every $f \in \Delta_T$:

$$P_T(f) = \frac{\sum_{j\leq qm} s_j f(j) P^j_{\max_j f(j)\cdot T}}{\sum_{j\leq m} s_j f(j)} \tag{9}$$

Proof

Obviously, we have to define $P_T^j = P_T(f^j)$ for all $T \geq 2$ and $j \leq m$.

For the representation, we have to show that there are positive numbers $(s_j)_{j\leq C}$ such that the representation holds for all $T \geq 2$ and for every $f \in \Delta_T$.

Rough sketch of the proof:

In general the proof follows the rough structure of BGSS, i.e. the idea to translate the framework form databases to frequencies to exploit the simplex structures. To derive the similarity values first for a set of basic cases consisting only of three cases and then use the gained results for the generalization to any finite number of basic cases is also based on BGSS. But except of the rough structure, the proof presented here needs different arguments to complete the different parts of the proof. In particular the anchored version version of the combination axiom requires a different way (compared to BGSS and hence also EG, which follows a very similar approach) to show the main crucial step of the proof. Namely, in BGSS the combination of any databases or frequency vector is allowed. Also in EG the combination of any frequency vectors (by taking care about the lengths and the constant similarity axiom) is basically possible. However in this paper only specific anchored databases or frequency vectors can be combined, which requires a different approach. As in EG, the constant similarity axiom is an important ingredient to facilitate the proof.

Rough Steps:

In Step 1 we proof the theorem for a set of basic cases consisting only of three different basic cases, i.e. $C = \{c_1, c_2, c_2\}$.

Step 1.1: Determination of the similarity values s_1, s_2, s_3

Similar to BGSS and EG, we derive the similarity weights $s_1, s_2, s_3 \in \mathbb{R}_+$, by applying the anchored concatenation axiom, constant similarity axiom and the diversity axiom. More specifically: The representation (8) in Lemma B.1 and the representation (9) in Theorem B.1 applied to $\overline{f} := \frac{1}{3}(f^1 + f^2 + f^3)$ yields (with the Diversity axiom) the similarity values, which allows the definition of $P^s_T(f) := \frac{\sum_{j \leq 3} s_j f(j) P^j_{\max_j f(j)T}}{\sum_{j \leq 3} s_j f(j)}$ for all $f \in \Delta(C)$ and $T \in \mathbb{N}$.

Step 1.2: Show that $P_T(f) = P^s_T(f)$ for all simplicial points (Figure 1 illustrates simplicial partitions and points)

The main tool to show this claim is the observation that for four specifically structured frequency vectors, which fulfill the above equation, also the intersection of the lines between two of these (specific) vectors satisfies the above equation (Lemma B.4). The crucial step in the proof is to apply his fact in a appropriate way (different than in BGSS, EG) inductively. In this step again the maximal anchored concatenation axiom and the constant similarity axiom (in form of Lemma B.2) are necessary.

Step 1.3: Show that $P_T(f) = P^s_T(f)$ for all frequency vectors $f \in \Delta(C)$

The proof is similar to a (rewritten/revised) proof of Lemma A.6 in EG, which is based on the existence of the limit of P^c_T for all $c \in C$ (Learning axiom). Since all frequency vectors $f \in \Delta$ can be approximated by a series of simplicical triangles/points, we can show the claim (by using Lemma B.1 and Lemma B.2). In particular, one can show that the beliefs P and P^s induced by the sequence of simplicial points, which approximates f, converges to the belief of P and P^s induced by the limit f. Using the equivalence of $P^s_T(g)$ and $P_T(g)$ for the sequence of simplicial points $g \in \Delta$ by Step 1.2 and the Diversity Axiom will deliver the claim.

In Step 2, the result from step 1 is used inductively for a general set of basic cases $C = \{c_1, c_2, ..., c_m\}$ with $m > 3$.

Step: 2.1: Defining the similarity weights $s_1, ..., s_m$

Step 1 is applied to any triple of cases $\{c_j, c_k, c_l\} \subseteq C$ for distinct $j, k, l \in \{1, 2, ..., m\}$, yielding similarity weights $s^{(j,k,l)}_j, s^{(j,k,l)}_k, s^{(j,k,l)}_l$. As in the proof of Proposition 3, Step 2.1 in BGSS, one can show that each similarity weight can be chosen independent of the choice of the triple, i.e. $s^{(j,k,l)}_j = s_j$. Hence as in Step 1.1. we can define $P^s_T(f) := \frac{\sum_{j \leq m} s_j f(j) P^j_{\max_j f(j)T}}{\sum_{j \leq m} s_j f(j)}$ for all $f \in \Delta(C)$ and $T \in \mathbb{N}$.

Step 2.2: Show $P_T(f) = P^s_T(f)$ for all $f \in \Delta(C)$

This is done inductively on $|M| = m$ for $f \in conv(\{(f^j)_{j \in M}\})$, where we use Step 1 for $m = 3$ as start of the induction. Each $f \in \Delta$ can be decomposed based on different anchors (Remark B.2). Applying the maximal anchored concatenation axiom to these decompositions yield hyperplanes, which are spanned by $(P(f^j_i(f(i))))_{j \neq i \leq m}$, for different $i \leq m$. All these hyperplanes contain $P(f)$ and also include $P^s(f)$ as well, since $P_T(f^j_i(k)) = P^s_T(f^j_i(k))$ for any $i \neq j \leq m$ and $f^j_i(k) \in \Delta_T$. Using the constant similarity axiom (Lemma B.2) and Lemma B.1 to harmonize the different hyperplanes wrt. lengths, we can show that the intersection of all these induced hyperplanes is unique, which delivers the desired result.

B.5 Step 1: $C = \{c_1, c_2, c_3\}$, i.e. $m = 3$

Step 1.1:

Define $\overline{f} := \sum_{j\leq 3} 1/3 f^j$, for $f^j \in \Delta_T$, and $T \geq T^*$ then $\overline{f} \in \Delta_{3T}$. We can choose positive numbers s_1, s_2, s_3 such that representation holds for $\overline{f}$ by equating the evaluation of $\overline{f}$ using the representation (8) in Lemma B.1 , i.e. $P_{3T}(\overline{f}) = \lambda_1 P_T^1 + \lambda_2 P_T^2 + (1-\lambda_1-\lambda_2)P_T^3$ with representation (9) in Theorem B.1 and solving the linear system. The solution of this linear system s_1, s_2, s_3 exist uniquely up to multiplication by a positive number due to the non collinearity condition of the Diversity Axiom for $T \geq T^*$, otherwise uniqueness is not achievable.
Define for all T and $f \in \Delta_T$

$$P_T^s(f) := \frac{\sum_{j\leq 3} s_j f(j) P^j_{\max_j f(j)T}}{\sum_{j\leq 3} s_j f(j)} \tag{10}$$

Obviously $P_T^s(f^j) = P_T(f^j)$ for all $j = 1, 2, 3$ and $P_T^s(\overline{f}) = P_T(\overline{f})$.
The aim is to show for all T and for every $f \in \Delta_T$:

$$P_T^s(f) = P_T(f) \tag{11}$$

In the following, we will partition the simplex Δ into so called simplicical triangles recursively, as illustrated in the Figure 1 below.

Definition of Simplicical Triangles:

The 0-th simplicial partition consist of vertices $q_0^j \in \Delta$, which are exactly the unit vectors f^j for $j = 1, 2, 3$. The first simplicial partition of Δ is a partition to four triangles separated by the segments connecting the middle points between the two of the three unit frequency vectors, i.e. $q_1^1 := (\frac{1}{2}f^1 + \frac{1}{2}f^2)$, $q_1^2 := (\frac{1}{2}f^2 + \frac{1}{2}f^3)$ and $q_1^3 := (\frac{1}{2}f^3 + \frac{1}{2}f^1)$. The second simplicial partition is obtained by similarly partitioning each of the four triangles to four smaller triangles, and the l-th simplicial partition is defined recursively. The simplicial points of the l-th simplicial partition are all the vertices of triangles of this partition. Note

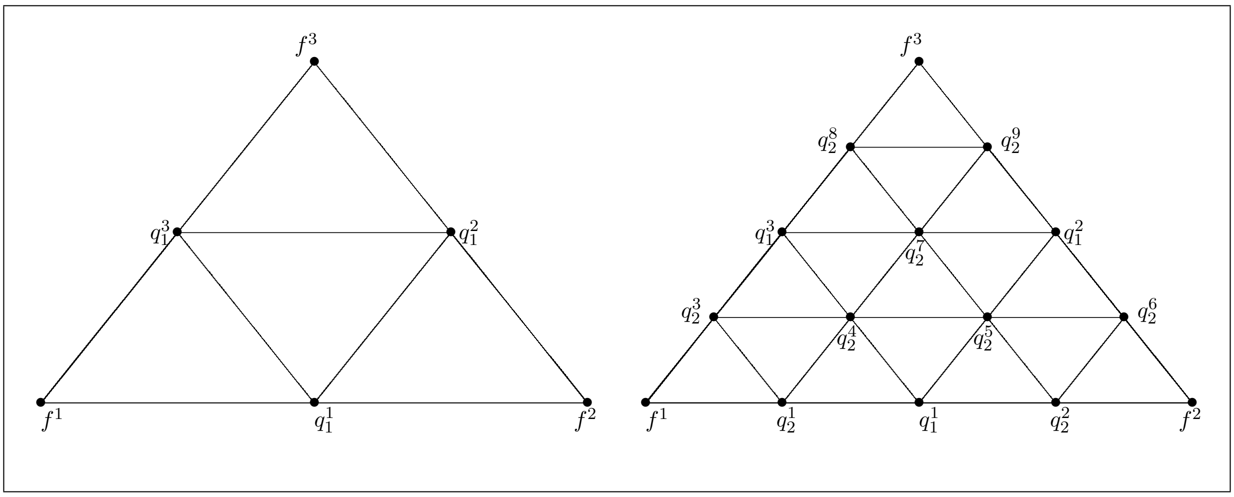

Figure 1: 1st and 2nd Simplicical partitions

that for $j = 1, 2, 3$ the q_0^j are frequency vectors representing databases consisting only of 1 case, but of any length $T \in \mathbb{N}$, i.e. $q_0^j \in \Delta_T$ for all $T \in \mathbb{N}_+$. All vertices q_l^v of the l-th simplicial partition are in $\Delta_{2^l T}$ for all $T \in \mathbb{N}_+$ for appropriate $v \leq n_l$ as defined below. Considering the simplicial points on the line between f^1 and f^2, we get for the 0-th simplicical partition: 2 simplicical points, for 1-th simplicical partition: 3 simplicical points: 3, for 2-th simplicical partition: simplicical points, for 3-th simplicical partition: 9 simplicical points and so forth, i.e. it follows the series $a_l = 2^l + 1$ for all $l \in \mathbb{N}$. Observe that for each parallel line to (f^1, f^2) between simplicial points of the l-th simplicical partition, the line which is one 'step' closer to f^3, possesses one simplicical point less than the farther parallel line. The number of simplicial points on these parallel lines decreases until reaching the point f^3. Hence the total number n_l of simplicial points of the l-th partition is given by

$$n_l := \sum_{i=1}^{a_l} i = \sum_{i=1}^{2^l+1} i = 2^{2l-1} + 2^l + 2^{l-1} + 1 \qquad \text{where } a_l = 2^l + 1 \tag{12}$$

$$\tag{13}$$

Step 1.2: Equation (11) holds for all Simplicial Points

In the following, we will partition the simplex into simplicical triangles and will show that the vertices of these triangles satisfies equation (11).

Lemma B.3
The vertices q_l^v with $v \leq n_l$ of the l-th simplicial partition satisfy equation (11) for all $l \in \mathbb{N}$.

Proof

Main tool of the proof is the following Lemma.
Notation: In the following we will denote for $a, b \in \Delta$ or $a, b \in \Delta(R)$ the straight line through a and b by (a, b) (since there won't be a confusion to the usual interval notation).

Lemma B.4
Let $a, b, c, d \in \Delta$ be distinct frequency vector satisfying equation (11) and the lines (a, b) and (c, d) are not collinear. Then the intersection y of the line (a,b) and (c,d), i.e. $y = (a, b) \cap (c, d)$ satisfies equation (11) (for an appropriate length T, i.e. such that $y \in \Delta_T$)) if the following conditions hold for both of the pairs a, b and c, d:
(i) both vectors a and b (respectively c and d) lie on a line $(f_i^j(k), f_i^h(k))$ for some $k \in [0, 1]$ and distinct $i, j, h \leq m$, which represent anchored databases with identical anchor case $c_i \in C$ or
(ii) a, b (respectively c, d) lie on a line between $(f_i^j(1), f_j^h(1/2))$ for some distinct $i, j, h \leq m$.

Proof
We will show the situation, where both pairs a, b and c, d satisfy condition (i).
Assume that $a, b \in (f_i^j(k), f_i^h(k))$, hence also $y \in (f_i^j(k), f_i^h(k))$. Hence by Remark B.2 we know that there exist a decomposition of y via $(f_i^j(k), f_i^h(k))$, i.e. there exist some $\alpha \in (0, 1)$ and $Z^y \in \mathbb{N}$ such that $y = \alpha f_i^j(k) + (1 - \alpha) f_i^h(k) \in \Delta_{Z^y}$ with corresponding adjusted length $L^y := L(k, \alpha Z^y, (1 - \alpha) Z^y)$ such that $P_{Z^y}(y) \in \big(P_{L^y}(f_i^j(k)), P_{L^y}(f_i^h(k))\big)$.
Analogously, there exist some Z^x and L^x for all $x \in \{a, b\}$. Let $L := LCM(Z^y, Z^a, Z^b)$, then for all $v \in \{y, a, b\}$ the following holds

$$P_{\frac{Z^v L}{L^v}}(v) \in \big(P_L(f_i^j(k)), P_L(f_i^h(k))\big)$$

In particular $P_{\frac{Z^a L}{L^a}}(a)$ and $P_{\frac{Z^b L}{L^b}}(b)$ determine already the shape/slope of the line $\big(P_L(f_i^j(k)), P_L(f_i^h(k))\big)$ and hence $P_{\frac{Z^y L}{L^y}}(y) \in \Big(P_{\frac{Z^a L}{L^a}}(a), P_{\frac{Z^b L}{L^b}}(b)\Big)$.

The same derivation can be executed with P^s and results in $P^s_{\frac{Z^y L}{L^y}}(y) \in \left(P^s_{\frac{Z^a L}{L^a}}(a), P^s_{\frac{Z^b L}{L^b}}(b)\right)$ and since we know that $a, b \in \Delta$ satisfy (11), we get:

$$P_{\frac{Z^y L}{L^y}}(y), P^s_{\frac{Z^y L}{L^y}}(y) \in \left(P_{\frac{Z^a L}{L^a}}(a), P_{\frac{Z^b L}{L^b}}(b)\right)$$

The same procedure applied to pair c, d instead to a, b with $G := LCM(L^y, L^c, L^d)$ yields:

$$P_{\frac{Z^y G}{L^y}}(y), P^s_{\frac{Z^y G}{L^y}}(y) \in \left(P_{\frac{Z^c G}{L^c}}(c), P_{\frac{Z^b G}{L^c}}(c)\right)$$

Finding a common multiplier $J = LCM(L, G)$ will deliver the desired result, since

$$P_{\frac{Z^y J}{L^y}}(y), P^s_{\frac{Z^y J}{L^y}}(y) \in \left(P_{\frac{Z^a J}{L^a G}}(a), P_{\frac{Z^b J}{G L^b}}(b)\right) \cap \left(P_{\frac{Z^c J}{L L^c}}(c), P_{\frac{J Z^d}{L L^d}}(d)\right)$$

and the intersection is unique (otherwise this would be a contradiction to the diversity axiom). Hence $P^s_{Z^y \frac{J}{L^y}}(y) = P_{Z^y \frac{J}{L^y}}(y)$ and by Lemma B.2 $P_T(y) = P^s_T(y)$ for all T such that $y \in \Delta_T$.

The situation, in which one of the two pairs satisfies condition (i) and the other condition (ii) or both pairs fulfill condition (ii) can be shown analogously. □

The proof of the theorem is conducted by using the observation in Lemma B.4 inductively, as can be seen in the series of figures (Figures 2 and 3) below.

Proof by induction over the l-th partition:
For $l = 0$:
By Step 1.1, we know that for $(q^1 = q^1_0, q^2 = q^2_0, q^3 = q^3_0)$ the representation holds.

Induction step:
Let the claim be true for the l-th simplicical partition. For the $(l+1)$-th partition the following procedure will capture all simplicical points q^v_{l+1} for $v \leq n_{l+1}$.
The procedure, which can be understand easily in the series of Figures 2 to 4, is expressed in quite extensive notational effort below.

We need some definitions:

(i) We will denote by $g^{(i,j)}_l(d)$ the simplicical point in the l-th partition on the line (f^i, f^j) such that it is the d closest to f^i. More precisely, let for all $i \neq j \in \{1, 2, 3\}$ and for all $l \in \mathbb{N}$:

$$g^{(i,j)}_l(d) \in \{q^x_l \in (f^i, f^j) \,|\, \text{there exists (d-1) many distinct } q^t_l \in (f^i, f^j) \text{ for } t, x \leq n_l \text{ s. th. } \|f^i - q^t_{l+}\| < \|f^i - q^x_l\|\}$$

where $||\cdot||$ is the standard norm on $\mathbb{R}^{|R|}$.
(ii) We denote by $b^i_{l+1}(d)$ the simplical point of the $(l+1)$-th partition, which lies on the line (f^i, q^*) and the d- closest line in the $(l+1)$-th partition parallel to f^j, f^h for distinct i,j,k.
Lines between these points are essential to cover all simplicial points through intersections. More precisely, define for all $\in \mathbb{N}$ and all distinct $i, j, h \in \{1, 2, 3\}$:
$b^i_{l+1}(d) := (f^i, 1/2(f^j + f^h)) \cap (g^{(i,j)}_l(d), g^{(i,h)}_l(d))$ (remember $1/2(f^j + f^h) = f^h_j(1/2) = q^w_1$ for some $w \in 1, 2, 3$, i.e. $1/2(f^j + f^h)$) satisfy equation (11)).

Procedure:
(i) **For** $d = 1$**:**

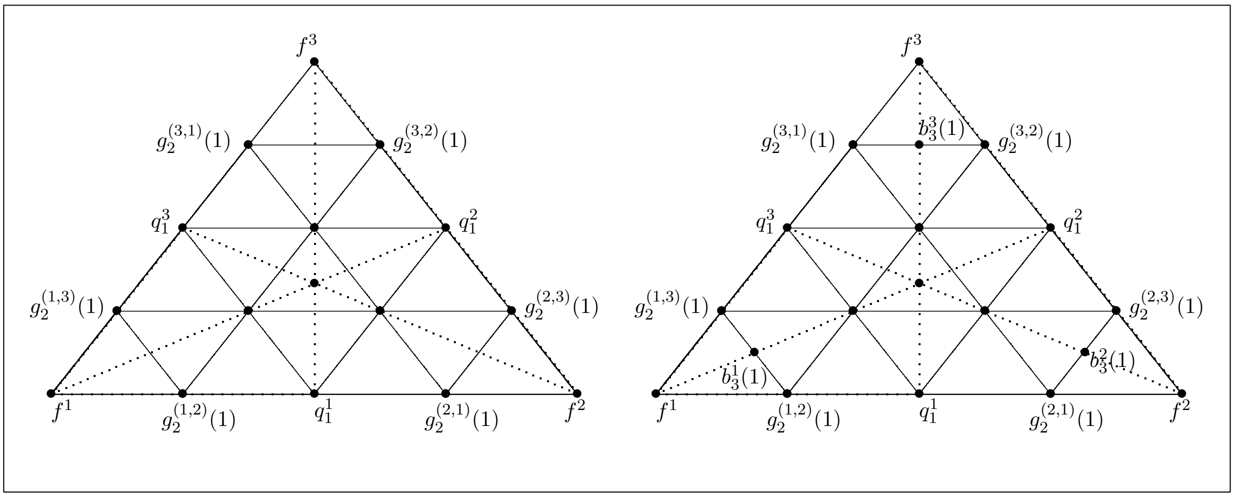

Figure 2: From 2nd to 3rd Simplicical partition points
Assume that all simplicial points (bullets) of the 2-nd partition satisfy already equation (11). Here some points are named according to the notation used in the procedure.
For example take the simplicial points of the 2nd partition that is on (f^1, f^2) and closest to f^1, i.e. $g_2^{(1,2)}(1)$. Analogously, take the closest to f^1 on (f^1, f^3), i.e. $g_2^{(1,3)}(1)$. Intersecting $(g_2^{(1,2)}(1), g_2^{(1,3)}(1))$ with (f^1, q_1^2) shows that $b_3^1(1)$ satisfies equation (11) as well. Analogously, this can be shown for $b_3^i(1)$ $(i = 2, 3)$ using appropriate combinations of $g_2^{(i,j)}(1), q_1^m, f^h$

W.l.o.g. take the perspective of $f^j = f^1$ for a $j \in \{1, 2, 3\}$. Given the l-th simplicical partition, there exist a simplicial point of the $(l+1)$-th simplicial partition $b_{l+1}^1(1)$, which is the intersection of the lines (f^1, f^2) and $(g_l^{(1,2)}(1), g_l^{(1,3)}(1)))$. By the induction assumption these pairs of points satisfy equation (11) and the conditions of the Lemma B.4 and hence $P_{2^{l+1}}^s(b_{l+1}^1) = P_{2^{l+1}}(b_{l+1}^1)$, i.e. $b_{l+1}^1(1)$ satisfies equation (11). Analogously the same procedure applied to f^j for $j = 2, 3$ yields that $b_{l+1}^j(1)$ satisfies equation (11).

(ii) Draw the line between two elements of $\{b_{l+1}^1(1), b_{l+1}^2(1), b_{l+1}^3(1)\}$, w.l.o.g. take $b_{l+1}^1(1)$ and $b_{l+1}^3(1)$. The line $(b_{l+1}^1(1), b_{l+1}^3(1))$ intersects for all $0 \le z \le a_l$ (defined above (12)) with the lines $(g_l^{(1,3)}(z), g_l^{(2,3)}(z))$ which are parallel to the line (f^1, f^2) and also with all lines $(g_l^{(1,2)}(z), g_l^{(1,3)}(z))$ which are parallel to (f^2, f^3). By Lemma B.4 this yields, that all simplicical points of the $(l+1)$-th partition, which lie on the line $(b_{l+1}^1(1), b_{l+1}^3(1))$ are satisfying equation (11).
Analogously, the procedure yields, that all simplicical points of the $(l+1)$-th partition, which lie on the lines $(b_{l+1}^i(1), b_{l+1}^j(1))$ for all combinations of $i \neq j \in \{1, 2, 3\}$ are satisfying equation (11), i.e. which are on the closest parallel lines to $(f^1, f^2), (f^1, f^3), (f^2, f^3)$ and in particular, the closest $(l+1)$-simplicial points to f^1, f^2, f^3 on the rim/boundary of $conv(\{f^1, f^2, f^3\})$.

(iii) Apply the procedure of (i) and (ii) (where $d = 1$) recursively for $d = 2n - 1 > 1$ for $2 \le n \le \frac{a_{l+1}-1}{2}$ (from f^1- view).
Derive $\{b_{l+1}^1(d), b_{l+1}^2(d), b_{l+1}^3(d)\}$ by (i) using f^h and $(g_l^{(i,j)}(d)$ for $i, j, h \in \{1, 2, 3\}$ appropriately. Using (ii), we can show that all simplicial points of the $(l+1)$-th partition, which lie on the lines $(b_{l+1}^i(d), b_{l+1}^j(d))$ for all combinations of $i \neq j \in \{1, 2, 3\}$ are satisfying equation (11).

Observe that for $d = 2n$ with $1 \le n \le \frac{a_{l+1}-1}{2}$ the simplicial points of the $(l+1)$-th partition, which lie on the lines $(b_{l+1}^i(d), b_{l+1}^j(d))$ for all combinations of $i \neq j \in \{1, 2, 3\}$ are already satisfying the equation (11) directly, since these lines already are 'used' for the

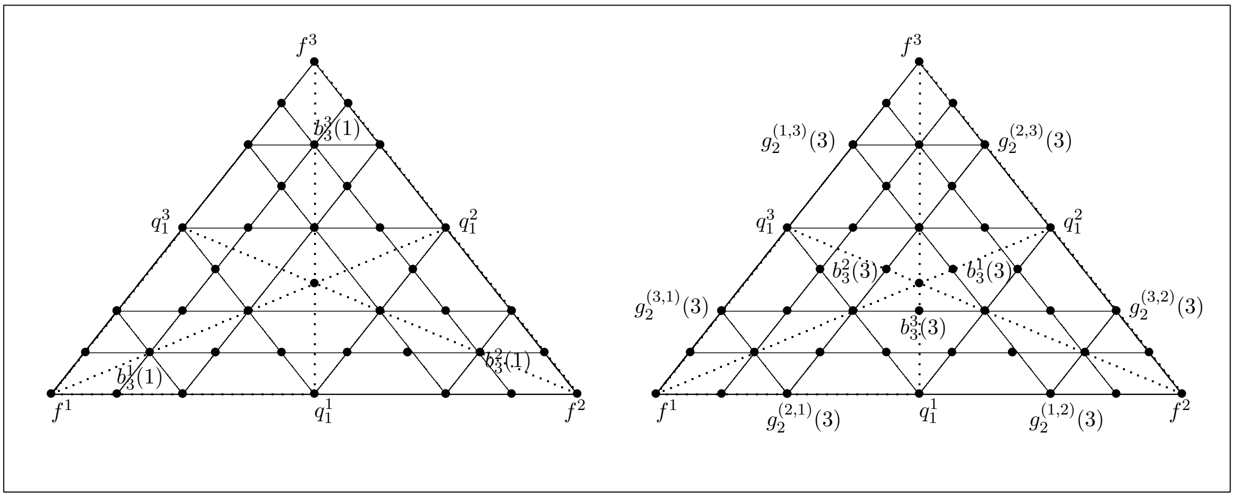

Figure 3: From 2nd to 3rd Simplicical partition points
a) Using lines $(b_3^i(1), b_3^j(1))$ and their intersections with existing lines, will show equation (11) for all simplicical points of the 3rd partition, which are on the closest parallel lines to the rim of the simplex.
b) Now, consider the simplicial points of the 2nd partition, which are on the lines (f^i, f^j) and third (since the second closest are covered indirectly) closest to f^i. For example, take $g_2^{(1,2)}(3)$ (third closest to f^1 on line (f^1, f^2)) and $g_2^{(1,3)}(3)$) (third closest to f^1 on (f^1, f^3)). Intersecting of $(g_2^{(1,2)}(3), g_2^{(1,3)}(3))$ with (f^1, q_1^2) shows that $b_3^1(3)$ satisfies equation (11) as well. Analogously, this can be shown for $b_3^i(3)$ $(i = 2, 3)$ using appropriate combinations of $g_2^{(i,j)}(3), q_1^m, f^h$.

procedure in the l-th partition and the simplicial points of the $(l + 1)$-th partition are just indirectly processed via the the intersection steps in (ii).

B.5.1 Step 1.3: Completion to all $f \in \Delta$

Lemma B.5
For all $2 \leq T \in \mathbb{N}$ *and* $f \in \Delta_T$*:* $P_T(f) = P_T^s(f)$*.*

Before we proof the Lemma, we mention some helpful considerations.
For each $f \in \Delta$ there exists a sequence of simplicical triangles $(q_l^{i_l}, q_l^{j_l}, q_l^{h_l})_{l \in \mathbb{N}}$ (remember $q_l^v \in \Delta_{2^l}$ for all $v \leq n_l$) for distinct $i_l, j_l, h_l \leq n_l$, such that:

(i) $f \in conv(\{q_l^{i_l}, q_l^{j_l}, q_l^{h_l}\})$ for all $l \in \mathbb{N}$, i.e. there exist $\beta_l^v \in [0, 1]$ for all $v \in \{i_l, j_l, h_l\}$ such that $f = \beta_l^{i_l} q_l^{i_l} + \beta_l^{j_l} q_l^{j_l} + \beta_l^{h_l} q_l^{h_l}$
(ii) For all $v \in \{i_l, j_l, h_l\}$ and $l \in \mathbb{N}$:
$q_l^v \in \Delta_{\beta_l^v H_l}$, such that H_l (as in Remark B.4) is the smallest common denominator of all β_l^v, i.e. there exist z_l^v, such that $\beta_l^v = \frac{z_l^v 2^l}{H_l}$. Hence, if f is represented by combination of an l-th simplicial triangle, then $f \in \Delta_{H_l}$
(iii) $lim_{l \to \infty} q_l^v = f$ for all $v \in \{i_l, j_l, h_l\}$

Clearly this construction is possible for all $f \in \Delta$.

To proof the Lemma, i.e. $f \in \Delta_T$: $P_T(f) = P_T^s(f)$, we proceed with the following steps. If we could show that $lim_{l \to \infty} ||P_{H_l}^s(f) - P_{H_l}^s(q_l^v)|| = 0$ and $\lim_{\to \infty} ||P_{H_l}(f) - P_{H_l}(q_l^v)|| = 0$ for all $v \in \{1, 2, 3\}$, the result would follow from $P_{H_l}(q_l^v) = P_{H_l}^s(q_l^v)$ (by Step 1.2) and the Diversity and Constant similarity axiom.

Proof of Lemma B.5
Step (i):

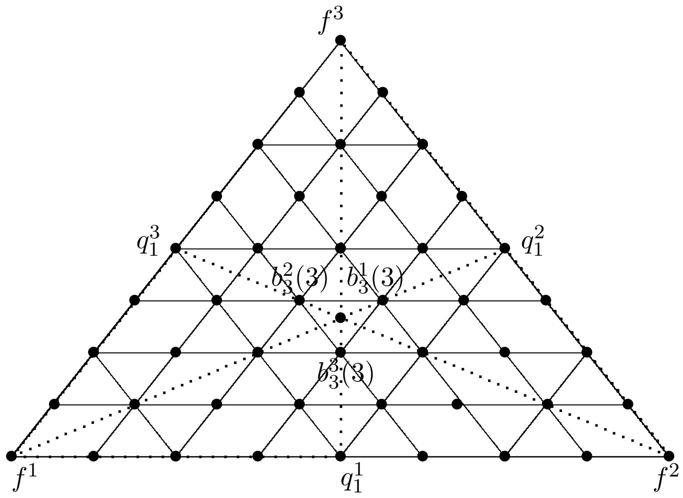

Figure 4: As before, the next step would be to intersect the lines between $(b_3^i(1), b_3^j(1))$ and the existing lines, will show equation (11) for all simplicical points of the 3rd partition, which are on the third closest parallel lines to the rim of the simplex, which completes the 3rd partition.

By the learning axiom and since $P_T^s(f^i) = P_T(f^i)$ for all $T \in \mathbb{N}$, we know that for all $i \leq 3$, we have $lim_{l\to\infty} P_{H_l}^s(f^i) = P_\infty(f^i) = P_\infty^i$.

We want to show for all $v \in \{i_l, j_l, h_l\}$: $lim_{l\to\infty} ||P_{H_l}^s(f) - P_{H_l}^s(q_l^v)|| = 0$.
Let for all $r \in R$, $P_T^i(r)$ be the r-th component of the probability vector.
For all l, $v \in \{i_l, j_l, h_l\}$ and q_l^v we have that $\lim_{l\to\infty} q_l^v = f$ and hence $\lim_{l\to\infty} P^j_{\max_j q_l^v(j)H_l}(r) = P^j_{\max_j f(j)H_l}(r)$ holds. This directly implies $\lim_{l\to\infty}(P_{H_l}^s(f)(r) - P_{H_l}^s(q_l^v)(r)) = 0$ for all $r \in R$ and hence the desired result.

Step (ii):

By Lemma B.1 we know, that $P_H(f) = \sum_{j=1}^3 \lambda_j P_{\max_{i=1,2,3} f(i)H}(f^j)$ where $\lambda \in \Delta^2$ is independent of the length of the database by the constant similarity axiom.
Hence by the learning axiom $lim_{l\to\infty} P_{H_l}(f) = lim_{l\to\infty} \sum_{j=1}^3 \lambda_j P_{\max_{\{i=1,2,3\}} f(i)(H_l)}(f^j)$ exists and hence with the same reasoning as above in the case of P^s, we know that

$$lim_{l\to\infty} ||P_{H_l}(q_l^v) - P_{H_l}(f)|| = 0 \quad \text{for all } v \in \{i_l, j_l, h_l\}$$

Step (iii):

By Step (i) and (ii) and the triangle inequality, we get:

$$lim_{l\to\infty} ||P_{H_l}^s(f) - P_{H_l}^s(q_l^v) - P_{H_l}(f) + P_{H_l}(q_l^v)|| = 0$$

Since for all l, we know that $P_{H_l}(q_l^v) = P_{H_l}^s(q_l^v)$, we have $lim_{l\to\infty}||(P_{H_l}^s(f) - (P_{H_l}(f))|| = 0$, which implies for all $r \in R$:

$$\begin{aligned} 0 &= lim_{l\to\infty}(P_{H_l}(f)(r) - P_{H_l}^s(f)(r)) \\ &= lim_{l\to\infty}\Big(\sum_{j=1}^{3}\lambda_j P_{\max_i f(j)H_l}(f^j)(r) - \frac{\sum_{j\le 3} s_j f(j) P^j_{\max_j f(j)H_l}(r)}{\sum_{j\le 3} s_j f(j)}\Big) \\ &= lim_{l\to\infty}\sum_{j=1}^{3} P_{\max_i f(j)H_l}(f^j)(r)\Big(\lambda_j - \frac{s_j f(j)}{\sum_{j\le 3} s_j f(j)}\Big) \\ &= \sum_{j=1}^{3} P_\infty(f^j)(r)\Big(\lambda_j - \frac{s_j f(j)}{\sum_{j\le 3} s_j f(j)}\Big) \end{aligned}$$

By the Diversity axiom, no three of $P_\infty(f^j)$ are collinear (i.e. also no $P_\infty(f^j)(r)$ are convex combinations), which implies that it must hold that $\lambda_j = \frac{s_j f(j)}{\sum_{j\le 3} s_j f(j)}$ for all $j = 1, 2, 3$, hence $P_{L_l}(f) = P_{L_l}^s(f)$ for all l and by the constant similarity Lemma B.2 $P_T(f) = P_T^s(f)$ such that $f \in \Delta_T$. □.

The proof for $C = \{c_1, c_2, c_3\}$ is concluded.

B.6 Step 2: $m > 3$:

B.6.1 Step 2.1 Defining the similarity weights:

Consider for $T \ge T^*$ and distinct $j, k, l \le m$ a triple $\{P_T^j, P_T^k, P_T^l\}$

Using the considerations from B.5 Step 1 for $\{j, k, l\}$, i.e. $\overline{f_{3T}} := \sum_{i\in\{j,k,l\}} f^i$ and $f^i \in \Delta_T$, we can derive the similarity weights $(s_i^{\{j,k,l\}})_{i\in\{j,k,l\}}$ and the following representation for all $f \in conv(\{f^j, f^k, f^l\}) \cap \Delta_T$:

$$P_T^{\{j,k,l\}}(f) = \frac{\sum_{i=j,k,l} s_i^{\{j,k,l\}} f(i) P^{\{j,k,l\}}_{max_i f(i)T}(f^i)}{\sum_{i=j,k,l} s_i^{\{j,k,l\} f(i)}}$$

Moreover for all $i \in \{j, k, l\}$, we have $P_T^{\{j,k,l\}}(f^i) = P_T(f^i) = P_T^i$ and $(s_i^{\{j,k,l\}})_{i\in\{j,k,l\}}$ are unique up to multiplication by a positive number.

Now we want to show, that the similarity values $s_i^{\{j,k,l\}}$ are independent of the choice of j, k and l for all $i \in \{j, k, l\}$. This can be shown in two steps:

1. Show that $\frac{s_j^{\{j,k,l\}}}{s_k^{\{j,k,l\}}} = \frac{s_j^{\{j,k,n\}}}{s_k^{\{j,k,n\}}}$,

i.e. the ratio between two similarity number is independent of the choice of a third case/frequency. Take two different triples $\{j, k, l\}$ and $\{j, k, n\}$, i.e. $l \ne n$. Consider the evaluation of rational combinations of $f^j \in \Delta_T$ and $f^k \in \Delta_T$, i.e. for $\alpha \in \mathbb{Q}$: $f = \alpha f^j + (1-\alpha) f^k$, where H is the smallest common denominator of $\alpha, (1-\alpha)$ and hence $f \in \Delta_H$, w.l.o.g. assume $\alpha \ge (1-\alpha)$. Then,

$$P_H^{\{j,k,l\}}(f) = \frac{s_j^{\{j,k,l\}}\alpha P^j_{\alpha H} + s_k^{\{j,k,l\}}(1-\alpha)P^k_{\alpha H}}{s_j^{\{j,k,l\}}\alpha + s_k^{\{j,k,l\}}(1-\alpha)} \text{ and } P_H^{\{j,k,n\}}(f) = \frac{s_j^{\{j,k,n\}}\alpha P^j_{\alpha H} + s_k^{\{j,k,n\}}(1-\alpha)P^k_{\alpha H}}{s_j^{\{j,k,n\}}\alpha + s_k^{\{j,k,n\}}(1-\alpha)}.$$

Equating these two expressions, we get:

$$\frac{s_j^{\{j,k,l\}}\alpha}{s_j^{\{j,k,l\}}\alpha + s_k^{\{j,k,l\}}(1-\alpha)} = \frac{s_j^{\{j,k,n\}}\alpha}{s_j^{\{j,k,n\}}\alpha + s_k^{\{j,k,n\}}(1-\alpha)} \text{ and } \frac{s_k^{\{j,k,l\}}(1-\alpha)}{s_j^{\{j,k,l\}}\alpha + s_k^{\{j,k,l\}}(1-\alpha)} = \frac{s_k^{\{j,k,n\}}(1-\alpha)}{s_j^{\{j,k,n\}}\alpha + s_k^{\{j,k,n\}}(1-\alpha)},$$

which leads to $\frac{s_j^{\{j,k,l\}}}{s_k^{\{j,k,l\}}} = \frac{s_j^{\{j,k,n\}}}{s_k^{\{j,k,n\}}}$

Denote this ratio by $S_{j,k} := \frac{s_j^{\{j,k,l\}}}{s_k^{\{j,k,l\}}}$, this ratio is defined for all distinct $j, k \leq m$, since strict positivity of the similarity numbers.
Further observe that the following holds:

$$S_{j,k}S_{k,l}S_{l,j} = \frac{s_j^{\{j,k,l\}}}{s_k^{\{j,k,l\}}} \frac{s_k^{\{j,k,l\}}}{s_l^{\{j,k,l\}}} \frac{s_l^{\{j,k,l\}}}{s_j^{\{j,k,l\}}} = 1 \tag{14}$$

2. Define $s_1 := 1$ and $s_j = S_{j,1}$ for all $j \leq m$.
Aim: To show that for all triple $j, k, l \leq m$ it holds that $s_i^{\{j,k,l\}} = as_i$ for some $a \in \mathbb{R}_+$.
If we can show that $\frac{s_i^{\{j,k,l\}}}{s_m^{\{j,k,l\}}} = \frac{s_i}{s_m}$ for for all $m \neq i \in \{j, k, l\}$, then we are done, since then $s_i^{\{j,k,l\}} = \frac{s_i}{s_m} s_m^{\{j,k,l\}} = as_i$ for all $m \neq i \in \{j, k, l\}$, e.g. with $m = k$ we have $a = \frac{s_k^{\{j,k,l\}}}{s_k}$ and hence $s_j^{\{j,k,l\}} = as_j$, $s_k^{\{j,k,l\}} = \frac{s_k^{\{j,k,l\}}}{s_k} s_k$, $s_l^{\{j,k,l\}} = as_l$.

hence it suffices to show w.l.o.g. that $\frac{s_j^{\{j,k,l\}}}{s_k^{\{j,k,l\}}} = \frac{s_j}{s_k}$ or equivalent $S_{j,k} = \frac{s_j}{s_k}$.
But the latter directly follows from (14), i.e. $1 = S_{1,j}S_{j,k}S_{k,1} = 1/s_j S_{j,k} s_k$, hence $S_{j,k} = \frac{s_j}{s_k}$ and hence the desired result.

The independence of the similarity values $s_i^{\{j,k,l\}}$ on $\{j, k, l\}$ allows to replace the (unique up to multiplication by a strictly positive number) $s_i^{\{j,k,l\}}$ by the just defined s_i for all $i \leq m$, i.e. given these $(s_i)_{i \leq m}$, one can define as in the consideration in B.5 Step 1.1:
For all $2 \leq T \in \mathbb{N}$ and any $f \in \Delta_T$.

$$P_T^s(f) := \frac{\sum_{i \leq m} s_i f(i) P^i_{max_i f(i) T}}{\sum_{i \leq m} s_i f(i)} \tag{15}$$

As in the section before the aim is to show that for all T and any $f \in \Delta_T$ the following equation holds: $P_T^s(f) = P_T(f)$.

B.6.2 Step 2.2: Completion to all $f \in \Delta$

Let $\Delta_T^M := \Delta_T \cap conv(\{f^j \mid j \in M\})$ denote the set of all frequency vectors $f \in \Delta_T$, which assign zero appearance to all cases $(c_i)_{i \in \{1,2,\ldots,m\} \setminus M}$, i.e. only cases $(c_j)_{j \in M}$ appear with positive frequency.
We will show by induction on $|M| = m$ for $3 \leq m \leq m$ the following claim.

Lemma B.6 *For every subset $M \subseteq \{1, 2, \ldots., m\}$ with $|M| = m \geq 3$, $P_T(f) = P_T^s(f)$, holds for every $f \in \Delta_T^M$*

Proof:

For $m = 3$ the claim has been shown in B.5 Step 1 (Step 1.3 in Lemma B.5).
Hence we assume now, that the claim holds for $m \geq 3$ and we prove it for M with $|M| = m + 1$.

1) Let $f \in \Delta_T^M$ such that $f \in conv(\{f^j\}_{j \in M \setminus l})$ for some $l \in M$, then by induction assumption $P_T^s(f) = P_T(f)$.
2) Now we consider $f \in int(conv(\{f^l \mid l \in M\}))$
By Remark B.2 we know that for all $i \in M$ and $l \in M \setminus \{i\}$, there exist for all $l \neq i \leq m$ some $\alpha_i^l \in (0, 1)$ and $\sum_{l \in M \setminus \{i\}} \alpha_i^l = 1$ such that $f = \sum_{l \in M \setminus \{i\}} \alpha_i^l f_i^l(f(i))$ with $f_i^l(f(i)) \in \Delta_{T_i^l}$ and then $f \in \Delta_{Z_l}$ (where $T_i^j = \alpha_i^j Z_i$). W.l.o.g. (due to constant similarity axiom, Lemma B.2) assume that for all $l \neq i \leq m$ there exist T_i^l such that $\max\{f(i), (1 - f(i))\} T_i^l \geq T^*$

(to overcome potential collinearity problems). For each $i \in M$ the corresponding adjusted lengths $L(f(i), (T_i^j)_{j \neq i \in M})$ are abbreviated by L_i in the following.
Now, the maximal anchored concatenation axiom induces that $P_{Z_l}(f)$ lies on the following induced $(m+1)$-many hyper-planes $A_l^{m+1}(Z_l)$ for all $l \in M$, w.l.o.g. assume that $M = \{1, 2, ..., m+1\}$:

$$
\begin{aligned}
P_{Z_1}(f) &\in int(conv(\{P_{L_1}(f_1^2(f(1))), P_{L_1}(f_1^3(f(1))), ..., P_{L_1}(f_1^{m+1}(f(1)))\})) =: A_1^{m+1}(Z_1) \\
P_{Z_2}(f) &\in int(conv(\{P_{L_2}(f_2^1(f(2))), P_{L_2}(f_2^3(f(2))), ..., P_{L_2}(f_2^{m+1}(f(2)))\})) =: A_2^{m+1}(Z_2) \\
&\in \cdots \\
P_{Z_{m+1}}(f) &\in int(conv(\{P_{L_{m+1}}(f_{m+1}^1(f(m+1))), P_{L_{m+1}}(f_{m+1}^2(f(m+1))), ... \\
&\qquad ....., P_{L_{m+1}}(f_{m+1}^m(f(m+1)))\})) =: A_{m+1}^{m+1}(Z_{m+1})
\end{aligned}
$$

Since for all $l \neq j \leq m$, $P_T^s(f_l^j(f(l))) = P_T(f_l^j(f(l)))$ for all T such that $f_l^j(f_l) \in \Delta_T$, we have also $P_{Z_l}^s(f) \in A_l^{m+1}(Z_l)$ for all $l \in M$.
For $Z = LCM(Z_1, ...Z_{m+1})$, the constant similarity axiom (Lemma B.2) implies that $P_Z(f), P_Z^s(f) \in A_l^{m+1}(Z)$ for all $l \in M$, i.e. $P_Z(f), P_Z^s(f) \in \bigcap_{l \in M} A_l^{m+1}(Z)$. By Lemma B.1 we have that for all $l \in M$ the sets $A_l^{m+1}(Z)$ consist of identical $(P^j_{max_{l \in M} f(l) Z})_{j \in M}$ (with different positive weights after evaluation of $P_Z(f_i^j(f(i)) = \lambda_j P^j_{max\{f(i),(1-f(i))\}Z} + (1 - \lambda_j) P^i_{max\{f(i),(1-f(i))\}Z}$ for particular $\lambda_j \in (0, 1))$.

This implies that determining $\bigcap_{l \in M} A_l^{m+1}(Z)$ means solving the $(m+1) \times (m+1)$ system of linear equations. We know that $|\bigcap_{l \in M} A_l^{m+1}(Z)| \geq 1$, since $P_Z^s(f)$ and $P_Z(f)$ are included in the intersection. The claim of $P_Z(f) = P_Z^s(f)$ would be proofed if we can show that $\bigcap_{l \in M} A_l^{m+1}(Z)$ is a singleton.
We will proof this by contradiction:
Assume that $P_Z(f) \neq P_Z^s(f)$, then the line $g := (P_Z(f), P_Z^s(f))$ has to be contained in $A_l^{m+1}(Z)$ for all $l \in M$. Hence this line g must intersect two of the faces H_j (of $dim(conv(\{P^j_{\max_{i \leq m} f(i) Z}\}_{\{j \in M\}})) - 1$), defined for all $j \in M$ by

$H_j := conv(\{(P^k_{\max_{i \leq m} f(i) Z})_{k \in M \setminus \{j\}}\})$. W.l.o.g. let these two faces be named H_u, H_v for some distinct $u, v \in M$. But then for all $l \in M$ $A_l^{m+1}(Z)$ has to intersect with these two faces H_u, H_v. We will show that this is not true. Observe that each $A_l^{m+1}(Z)$ intersects with all $(H_j)_{j \neq l \in M}$. Further, observe that applying the successive intersection, we get for $t \leq m+1$ $\cap_{j=1}^t A_j^{m+1}(Z) = \{H_{t+1}, ...H_{m+1}\}$, which implies $\cap_{j=1}^{m+1} A_j^{m+1}(Z) = \emptyset$, i.e. there exist no $dim(conv(\{P^j_{\max_{i \leq m} f(i) Z}\}_{\{j \in M\}})) - 1$)-faces such that all $A_l^{m+1}(Z)$ intersect them. Hence there cannot exist g such that $g \in A_l^{k+1}(Z)$ for all $l \in M$, which implies that there cannot be more than one unique element in the intersection of all $A_l^{m+1}(Z)$, i.e. $\cap_{l \in M} A_l^{m+1}(Z) = P_Z^s(f) = P_Z(f)$. By Lemma B.2 we get $P_T(f) = P_T^s(f)$ for all T such that $f \in \Delta_T$, which completes the proof of the Theorem B.1 and hence also Theorem 5.1.
□

C Incompatible objective belief formation

An objective belief without any imagination effort for each case reads

$$P_T(D) = \frac{\sum_{j \leq C} s(c_j) f_D(c_j) P^{c_j}_{f_D(c_j) \cdot T}}{\sum_{j \leq C} s(c_j) f_D(c_j)}$$

Is there a modification of the concatenation axiom (as stated in equation (3)), that is necessary for an objective representation and which database are admissible?

Modified Version of the Concatenation axiom in the sense of:

For two databases $D \in \mathcal{D}^{T_1}$ and $E \in \mathcal{D}^{T_2}$, $T := T_1 + T_2$ and two numbers $L, H \in \mathbb{N}_+$ such that $f_D \cdot L \in \mathbb{N}^m$ and $f_E \cdot H \in \mathbb{N}^m$, there exists a $\lambda \in (0,1)$ such that

$$P_T(D \circ E) = \lambda P_L(D^{L/T_1}) + (1-\lambda) P_H(E^{H/T_2}).$$

W.l.o.g. we will restrict the analysis to a set of three basic cases, i.e. $C = \{c_1, c_2, c_3\}$. Applying the objective belief to the modified concatenation axiom yields:

$$\begin{aligned}\frac{\sum_{j\le 3} s(c_j) f_{D\circ E}(c_j) P^{c_j}_{f_D(c_j)T_1 + f_E(c_j)T_2}}{\sum_{j\le 3} s(c_j) f_{D\circ E}(c_j)} &= \lambda \frac{\sum_{j\le 3} s(c_j) f_D(c_j) P^{c_j}_{f_D(c_j)L}}{\sum_{j\le 3} s(c_j) f_D(c_j)} \\ &\quad + (1-\lambda) \frac{\sum_{j\le 3} s(c_j) f_E(c_j) P^{c_j}_{f_E(c_j)H}}{\sum_{j\le 3} s(c_j) f_E(c_j)}\end{aligned}$$

We are not allowing to have a law of dynamics for the probabilities $P_T^{c_j}$ (which is also not reasonable), i.e. like some function Y of $P_T^{c_j} = Y(P_L^{c_j}, P_H^{c_j})$. Thus, we directly need to equalize the precision level for the estimations for each single case, i.e. for all $j \le m$

$$1_{\{f_D(c_j)>0\}} f_D(c_j) L = 1_{\{f_E(c_j)>0\}} f_E(c_j) H \text{ and } f_D(c_j) L \in \{0, f_D(c_j)T_1 + f_E(c_j)T_2\} \quad (16)$$

Let $i, j, l \in \{1, 2, 3, \}$ be mutually distinct.

In **Situation 1** let there exist some $i \le m$ such that $f_D(c_i) = 0$. Consequently, we have $f_{D\circ E}(c_i)T = \frac{f_E(c_i)T_2}{T} T$, which must be equal to $f_E(c_1)H$ and implies $H = T_2$. Only disjoint D and E can satisfy (16) with $H = T_2$. Thus databases D with $f_D(c) = 0$ for some $c \in D$ allow only concatenations of disjoint databases.

Hence in **Situation 2** we consider only databases that share the same support, i.e. $f_D(c) > 0$ iff $f_E(c) > 0$. For $m = i, j$ equations (16) can be rewritten to

$$L = \frac{f_D(c_m)T_1 + f_E(c_m)T_2}{f_D(c_m)} = T_1 + \frac{f_E(c_m)T_2}{f_D(c_m)} \text{ and } H = T_2 + \frac{f_D(c_m)T_1}{f_E(c_m)}.$$

The first equation implies that $L = T_1 + \frac{f_E(c_i)T_2}{f_D(c_i)} = T_1 + \frac{f_E(c_j)T_2}{f_D(c_j)}$, which results in $\frac{f_E(c_i)}{f_D(c_i)} = \frac{f_E(c_j)}{f_D(c_j)}$, which is only feasible if $\frac{f_E(c)}{f_D(c)} = 1$ for all $c \in D$. Hence $f_D = f_E$ represent different replications of the same database and the above equations yield $L = T_1 + T_2 = H$. Thus, the only non-disjoint concatenating database a modified concatenation axiom allows for are replicated identical databases, which is naturally true for all $\lambda \in (0,1)$, i.e. $P_{T_1+T_2}(A) = \lambda P_{T_1+T_2}(A) + (1-\lambda) P_{T_1+T_2}(A)$.

In sum, a objective belief satisfies a modified concatenation axiom only for combinations of disjoint databases or replicated identical databases. However, as seen in the proof a restriction to disjoint combining databases offers not sufficient structure to combine all frequency vectors in the simplex, i.e. the axiom allows only for combination of some lower dimensional hyper-planes, and hence will not allow to derive the desired objective belief.

D Relationship to EG's Axiom "Concatenation restricted to databases of equal length"

As mentioned at the end of Section 4.2.2, for a general concatenation axiom, the immense compromising considerations between different cautious estimations can only be avoided by assuming a common arbitrary level of precision according to which all cases are estimated, independent of the objective precision of the information. For each piece of information, literally agents need to imagine (or forget) sufficiently many observation of cases to reach an assumed artificial common level of precision. This ensures that no considerations and compromising regarding different precisions is required and allows an easy averaging based only on relative relevances of the concatenating databases. However, thereby an agent also needs to know a priori that she evaluates all information in an imagined precision and the beliefs contains only (imagined) equally precise and cautious estimations. Consequently, this means that a version of a concatenation axiom that cares for precision and also applies to arbitrary non-disjoint concatenations accomplishes the averaging of differently precise information by explicitly assuming away the differences in the sense of employing consciously estimations induced by imagined (forgotten) equally precise information.

Nevertheless, this discussion delivers an explanation and intuition for the (unexplained) statement in EG: " ... we modify the concatenation axiom of BGSS by restricting it to databases of equal length, i.e. thus controlling for the ambiguity resulting from insufficient amount of data ". Their restriction to equal lengths is ad hoc. However, technically one could argue for the equal length assumption by referring to the discussion above. An aggregation of differently precise information is only feasible if estimations are based on a common precision level, which is a consequence of the restrictions in their axiom. More detailed, their axiom demands, that for a set of n databases of the same length T, that can be concatenated to a n-times replication of a database, a belief induced by this database (not the n-th replication) is a average of the beliefs induced by each of the n databases separately. Obviously, this implies for a appropriate set of concatenating databases (consisting only of a single case) [24] that the belief induced by the database - which underlie the n-th replication- is formed by an average of the beliefs induced by T-times observed cases, i.e. for some appropriate $(\lambda_c)_{c\in D} \in (0,1)$

$$P(D) = \sum_{c\in D} \lambda_c P(c^T)$$

Thus, the restriction to equal lengths implies directly that all contained estimation are based on this common level as well. In this way EG is indirectly exploiting the above mechanism to overcome the immense compromising considerations.

However as already discussed, from our perspective and motivation, equal lengths of database are not sufficient to control for precision of its contained information. Moreover, in the spirit of the above discussion, EG's restriction to equal lengths cannot be meaningful interpreted as controlling for imprecision, but more as an (implicit) proposal to employ the length of the entire databases as the common (imagined) precision level according to which all estimations are made.

E Minimal Anchored Axiomatization

Instead of focussing on the most precise case in a database to determine the precision of its induced belief, we can also take the least precise case as the key determinant for the precision of a belief. This modification results in a minimal version of an anchored Concatenation and Constant Similarity Axiom and a corresponding extremely cautious belief formation.

Definition E.1
Let $F \in C^T$ be an anchored chain of $(D_i^j(k,T_j))_{j\neq i\leq m}$.
A length $M \in \mathbb{N}$ is called the ***adjusted (minimal) length*** *and denoted by $M(k,(T_j)_{j\neq i\leq m})$ if it is such that the number of observations of the least frequent case in an anchored chain $F \in C^T$ is identical to the number of observations of the least frequent case in the anchored databases $D_i^j(k,L)$ (for all $j\neq i$), (i.e. $\min_{c\in C} f_F(c)T = \min_{c\in C} f_{D_i^j(k,M)}(c)M$)*

[24]This is always possible, for example consider $D = (c_1^2, c_2, c_3^3) \in C^6$, then $D^6 = (c_1^6)\circ(c_1^6)\circ(c_2^6)\circ(c_3^6)\circ(c_3^6)\circ(c_3^6)$, which implies $P(D) = \lambda_1 P(c_1^6) + \lambda_2 P(c_2^6) + (1-\lambda_1-\lambda_2)P(c_3^6)$

Minimal Anchored Concatenation Axiom:

(i) Let $F \in C^T$ be an anchored chain of $(D_i^j(k, T_j))_{j \neq i \leq m}$, i.e. $F = \circ_{j \neq i}^m D_i^j(k, T_j)$ and let $M \in \mathbb{N}$ be the corresponding adjusted (minimal) length, i.e. $M = M(k, (T_i^j)_{j \neq i})$, then there exists $\lambda \in \Delta^m$ (where $\lambda_j = 0$ for all $j \leq m$ s. th. $T_j = 0$), s. th.

$$P_T(F) = \sum_{j \neq i \leq m} \lambda_j P_M(D_i^j(k))$$

(ii) Let for three distinct $i, j, l \leq m$ and any $V, W \in \mathbb{N}$ $F = D_i^j(1, V) \circ D_j^l(1/2, 2W)$ then there exist $\lambda \in int(\Delta^2)$:

$$P_{V+2W}(F) = \lambda P_{\min\{V,W\}}(D_i^j(1)) + (1 - \lambda) P_{\min\{2V,2W\}}(D_j^l(1/2))$$

For an analogously adjusted Constant Similarity Axiom the resulting theorem reads:

Theorem E.1
Let there be given a function $P : C^* \to \Delta(R)$. Let P_T be the restriction of P to C^T for $T \in \mathbb{N}_+$. Let P satisfies the Learning Axiom and the Diversity Axiom.
Then the following are equivalent:

(i) The function P satisfies the Invariance axiom, the minimal anchored Concatenation axiom, the (minimal) Constant Similarity axiom
(ii) There exists for each $(T, c) \in \mathbb{N} \times C$ a unique $P_T^c \in \Delta(R)$, and a unique -up to multiplication by a positive number- function $s : C \to \mathbb{R}_+$, s. th. for all T and any $D \in C^T$:

$$P_T(D) = \frac{\sum_{c \in D} s(c) f_D(c) P^c_{T^D_*}}{\sum_{c \in D} s(c) f_D(c)} \tag{17}$$

where $T_D^* \in \mathbb{N}_+$ is defined by $T_*^D := T \cdot \min_{c \in D} f_D(c)$.

Interpretational, this means that all estimations are based on the least precise information contained in the database and no information needs to be imagined. However, the focus on the least precise information results in neglecting and discarding many more precise pieces of information, by processing only until the level of least precision.

A detailed interpretation in terms of perception of precision and an adoption of an implied attitude of extreme cautiousness can be found in the discussion after Theorem 5.1.

References

[1] Ahn, D. (2008). Ambiguity Without a State Space. *Review of Economic Studies, 71, 3-28.*

[2] Arad, A., Gayer, G. (2012). Imprecise Data Sets as a Source of Ambiguity: A Model and Experimental Evidence. *Management Science, 58-1, 188-202.*

[3] Billot, A., Gilboa, I., Schmeidler, D. (2004). An Axiomatization of an Exponential Similarity Function. *Mimeo, University of Tel Aviv.*

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

[5] Bleile, J. (2014). Limited attention in Case-based Belief Formation. *Unpublished Paper.*

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

[7] Ellsberg, D. (1961). Risk, ambiguity and the Savage axioms. *Quartely Journal of Economics, 75, 643-669.*

[8] Epstein, L.G., Schneider, M. (2007). Learning Under Ambiguity. *Review of Economic Studies, 74, 1275-1303.*

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

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

[11] Gilboa, I., Schmeidler, D. (2003). Inductive Inference: An Axiomatic Approach. *Econometrica, 71, 1-26.*

[12] Gilboa, I., Lieberman, O., Schmeidler, D. (2006). Empirical Similarity. *Review of Economics and Statistics, 88, 433-444.*

[13] Gilboa, I. (2009). Theory of Decision under Uncertainty. *Econometric Society Monographs. Cambridge University Press, Cambridge.*

[14] Gilboa, I., Schmeidler, D. (2010). Case based Predictions: Introduction.

[15] Gilboa, I., Lieberman, O., Schmeidler, D. (2011). A similarity-based approach to prediction. *Journal of Econometrics, 162-1, 124-131.*

[16] Gilboa I., Postlewaite, A., Schmeidler, D. (2012). Rationality of belief or: why savages axioms are neither necessary nor sufficient for rationality. *Synthese, 187-1, 11-31.*

[17] Goldstone, R.L., Son, J.Y. (2005). Similarity. *The Cambridge handbook of thinking and reasoning. Cambridge University Press, 13-36.*

[18] Hau, R., Pleskac, T.J., Hertwig, R. (2010). Decisions From Experience and Statistical Probabilities: Why They Trigger Different Choices Than a Priori Probabilities *Journal of Behavioral Decision Making, 23, 48-68.*

[19] Heit, E. . Features of Similarity and Category-Based Induction. *Proceedings of the Interdisciplinary Workshop on Categorization and Similarity, University of Edinburgh, 115-121.*

[20] Heit, E. and Rubinstein, J. (1994). Similarity and Property Effects in Inductive Reasoning. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20-2, 411-422.*

[21] Marinacci, M. (2002). Learning about Ambiguous Urns. *Statistical Papers, 43, 143-151.*

[22] Rubinstein, A., Salant, Y. (2006). A model of choice from lists *Theoretical Economics, 1, 3-17.*

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

[24] Schank, R. C. (1986). Explanation Patterns: Understanding Mechanically and Creatively. *Hillsdale, NJ: Lawrence Erlbaum Associates.*

[25] Tversky, A. (1977). Features of Similarity. *Psychological Review, 84-4, 327-352.*