Limited Attention in Case-Based Belief Formations

Jörg Bleile
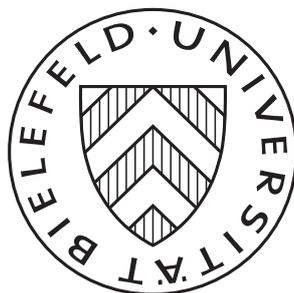
Limited Attention in

Case-Based Belief Formation

Jörg Bleile[*]

June 30, 2014

Abstract

An agent wants to derive her belief over outcomes based on past observations collected in her database (memory). There is well establish evidence in the psychology and marketing literature that agents consistently fail (or choose not) to process all available information. An agent might be constraint to pay attention (recall) and consider only parts of her potentially available information due to unawareness, cognitive or psychological limitations or intentionally for effort-efficiency. Based on this insight, we axiomatize a two-stage belief formation process in which in a first step agents filter ((un)intentionally) the available information. In a second step individuals employ the remaining observations to express a belief. We impose cognitively and normatively desirable properties on the filtering process. The axioms on the belief formation stage describe the relationship between databases and their induced beliefs. The axiomatized belief induced by a filtered databases is representable by a similarity weighted average of the estimations induced by each past attention-grabbing observation. An appealing application is a satisficing filter that induces a filtered belief that relies only on past experiences that are sufficiently relevant for a current problem. For the specific situation that agents (are able to) always pay attention to all available information, our axiomatization coincides with the axiomatization of a belief formation in Billot et al. (Econometrica (2005)).

Keywords: Belief formation, prior, case-based reasoning, relative frequencies, similarity, limited attention, consideration set, heuristics, satisficing, multicriteria choice, rationalization

JEL Classification: D01, D03, D11 (D81, D83)

1 Introduction and Motivation

In many situations agents need to evaluate uncertain consequences of their actions. In order to compare different potential consequences agents need to assign likelihoods to these outcomes. How can individuals form (probabilistic) beliefs over outcomes?

Traditionally, economic theory models uncertainties in a state space representation a la Savage (1954) and Bayes and derive a subjective prior based on observable actions of the agent. This implicitly requires that an agent already posses a subjective prior belief, which is expressed by her observable actions. However, the Savage and Bayesian approach does not help an agent to find or form a prior explicitly, for instance by incorporating pieces of information directly into a belief formation. In particular in situations in which an agent might not be able to condense her insufficient or too complex information into a consistent state space, their normatively appealing and convincing approach to endogenously derive a belief is not feasible [1].

We will consider such an environment and axiomatize a belief formation that allows to take directly into account the available information. This is strongly related to the aim of (asymptotic) statistical inference, where from data a distribution is derived. However in this paper we give a behavioral foundation for a belief formation in "non-asymptotic environments" that are characterized by heterogenous and limited information gathered in a list or database.

The impact of data and experience on the formation of a probabilistic belief was examined initially by the axiomatization of Billot et al. (2005) (BGSS from now on). The axiomatizations of BGSS and related ones of Eichberger and Guerdjikova (2010) (EG) (for ambiguous multiprior beliefs) and Bleile (2014a) (B) (precision dependent cautious beliefs) yield that a belief induced by a database is a similarity weighted average of the estimations induced by **all** observed cases in the database [2]. Thereby similarity weights capture different degrees of relevance of the potentially very heterogenous information [3].

A common shortcoming of these approaches is that an agent is obliged to take into consideration and account **all** past observations in her database. This precludes reasonable situations in which an agent might want to neglect, does miss or just forgets some pieces of information that would be in principle available. Our work relaxes this drawback of "compulsory" paying attention to all obtainable information. For this purpose we extend the mentioned axiomatic approaches (in particular BGSS) by adding a component of limited attention or consideration regarding available information.

A traditional and widely accepted assumption in economic theory is that gaining more information is beneficial and leads to improved actions. In this way, it is usually assumed that agents incorporate and take into account all available pieces of information [4]. However, the assumption of full attention and consideration of all available information requires that agents are aware of it, perceive it (unbiased) and eventually are able to process it without

---

[1]For more details on the difficulties of the Savage and Bayesian approach, see e.g. Gilboa et al. (2012) and Bleile (2014a)

[2]The framework is based on Case based Decision Theory and its application to prediction problems (Gilboa and Schmeidler (1995, 2001, 2003)).

[3]Basically, the three axiomatizations differ in the way the estimations are made.

[4]In addition, it is also a key assumption in all traditional revealed preference approaches.

any cognitive and psychological constraints [5].

The idea and concept of limited attention goes back to the seminal studies in psychology of Miller (1956), in which he identified limited cognitive abilities in processing information as the source of incomplete consideration, especially deficits and constraints in parallel (simultaneous) processing of information. Since then, mounting evidence in psychology and marketing shows that agents process and restrict attention to only a small fraction of the overall available information and consistently fail to consider all potentially available information due to their limited attention span (e.g. Broadbent (1958), Stigler (1961), Pessemier (1978), Hauser and Wernerfelt (1990), Chiang et al. (1998)). Often agents employ (implicitly) a multistage process to assign different degrees of attention to specific pieces of information (Bettman (1979)). In an initial rough filtering or screening stage agents pre-selects these elements that receive (or are worth to capture) full attention and consideration. In the literature this set of "surviving" elements is called a consideration set (Wright and Barbour (1977), Bettmann (1979), Roberts and Lattin (1991)).

A formation of a consideration set might emerge for many reasons. Cognitive constraints in parallel processing of information and unawareness of the presence of information (due to complexity, size, sequential processing or search) might cause an unintentional filtering (Miller (1956), Nedungadi (1990), Schwartz (2004)). The formation of a consideration sets as a (unintentionally) reply to avoid cognitive overload has been also studied in economic problems, e.g. recently Masatlioglu et al. (2012) axiomatized choices under (unintentional) limited attention.

In contrast, a consideration set can also be the result of a purposeful strategic elimination process. Agents often use (heuristic) filtering procedures to screen information rapidly and roughly before engaging into a costly and detailed evaluation (e.g. Wright and Barbour (1977), Gensch (1987), Nedungadi (1990), Gigerenzer et al. (1999), Hauser (2013)). Usually, these heuristics are noncompensatory cutoff or satisficing rules that allow for an uncomplicated "effort-efficient" comparison. This approach has recently gained prominence in economics (in particular in decision theory), e.g. through the works of Lleras et al. (2010) and Eliaz and Spiegler (2011a,b).

Another reason for the emergence of a consideration set relies on mounting evidence from psychology showing that often non-objective criteria like value systems, subjective motives or aversions, etc. restrict the attention of agents. Recent work has modeled these subjective and psychological biases ranging from overwhelming temptation (Gul and Pesendorfer (2001), Dekel and Lipman (2012)), rationalization (Cherepanov et al. (2013)), status quo bias (Masatlioglu and Ok (2005)), routes (Apesteguia and Ballester (2013)) and reason based choice (Lombardi (2009), De Clippel and Eliaz (2012)).

In this paper, we want to incorporate the formation of a consideration set induced by limited attention (consideration) as an intermediate stage into a belief formation process. In this way, our agent is not obliged to take into account all potentially available information, but might base her belief only on (survived) filtered information in the consideration set. In order to illustrate the basic idea and plausibility of such a two stage belief formation process we modify the doctor example of BGSS.

---

[5]According to Simon (1959, p.272) perception and cognition intervene between subjective view and the objective real world. In this context perception is often referred to as a "filter", where filtering can not only be seen as a passive, but also as an active selection process involving exclusion of almost all that is not within the scope of attention.

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

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

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

$$\text{"BGSS-belief":} \qquad P_x(D) = \frac{\sum_{c \in D} s(x, x_c) P^c}{\sum_{c \in D} s(x, x_c)} \qquad (1)$$

However, if the database D is long, complex or retrieved partly from her memory, the doctor might not want or is just not able to (recall) pay attention to and take into account all potential cases in the database D. She filters out some observations contained in D with specific features $\Gamma(D) \subseteq D$. An intuitive example is a **similarity satisficing** procedure, in which she considers only sufficiently relevant cases that surpass a threshold of similarity s^*, i.e. $\Gamma(D) = (c \in D | s(x, x_c) \geq s^*)$:

$$\text{"Filtered-belief":} \qquad (P_x \circ \Gamma)(D) = \frac{\sum_{c \in D} \mathbf{1}_{\{s(x,x_c) \geq s^*\}} s(x, x_c) P^c}{\sum_{c \in D} \mathbf{1}_{\{s(x,x_c) \geq s^*\}} s(x, x_c)} \qquad (2)$$

The filtered belief formation based on the similarity satisficing principle (under additional restrictions on the threshold value s^*) represents a special case of the general result we obtain in our representation theorem in Section 5.1.

Roughly, our filtered belief formation consist of two stages, in which initially a subjective (specific) filter process "selects" the information that builds the consideration set. In the second step, the agent forms her belief based on the remaining un-eliminated information in her consideration set.

One might be tempted to interpret such a filtered belief as the belief (1) of BGSS, based on an already (exogenously) ex-ante and independently filtered database $\Gamma(D)$. However, such an separation of filtering and belief formation would exclude plausible and appealing filters based on similarities (as in (2)), since the similarity values are endogenously derived in the belief formation. Moreover, in our axiomatization both stages are merged by an axiom that focusses on the relationship between filtered databases and their induced beliefs.

The initial filtering process captures appealing and desirable psychological properties

that are rooted in psychology and marketing literature. The main property is the well known and accepted consideration property. It is based on the idea and evidence, that if a case is considered in a database, i.e. is attention grabbing, then it should attract attention also in all of its sub-databases, since it faces less competition for attention by fewer pieces of information [6]. Further, we make some assumptions on the cognitive ability of agents and assume that an agent is able to pay attention to at least k ($k \geq 3$)-many available different pieces of information. A slightly more demanding characteristics requires that order and frequency in which information appears in a databases does not affect the level of attention an agent attributes to it. Basically that means that pieces of information are per se attention grabbing and not due to their specific position or a sufficiently high number of appearance.

The second layer of a filtered belief formation concerns the axiomatization on the belief level. The normatively reasonable axioms follow the basic intuition of the axioms in BGSS, but are modified to capture the previous filtering stage. We generalize the Concatenation Axiom of BGSS in order to capture the previous filtering process on the involved databases. The original Concatenation Axiom says that a belief induced by a combination of two databases is formed as an average of the beliefs that are induced by each of these databases separately. We cannot directly translate this to filtered database, since concatenations of already filtered databases can differ enormously from the result of filtering the concatenation of the two underlying databases (before they were filtered)[7]. Thus, in order to ensure a reasonable averaging of the induced beliefs in the spirit of the axiom, we require consistent relationships between the involved filtered databases. Another (implicit enforced) axiom ensures that the order in which information appears is irrelevant for the belief it induces.

As a result, our filtered belief formation can be represented as a similarity weighted average of the estimates induced by each case that the agent actually pays attention to, i.e. of those that survive the filtering. Hence the representation coincides with BGSS if the agent does not filter any information, but takes into account all available information.

Apart from the appealing intuition of filtering according to similarities (as in (2)), the filtering process can be any general arbitrary process that satisfies the required properties. Various subjective and psychological motives, constraints, biases and justifications can be employed as elimination criteria [8]. In particular, many recently developed multi-criteria decision procedures include elimination procedures to (implicitly) form a consideration set. The literature varies in the kind of criteria that are employed, e.g. using (compositions of) rationale(s) (sequentially) to eliminate alternatives (Manzini and Mariotti (2007, 2012a), Apesteguia and Ballester (2013), Houy (2007, 2010), Houy and Tadenuma (2009), Horan (2013)), focusing only on subjectively justifiable alternatives (Cherepanov et al. (2013), Gerasimou (2013)), considering only alternatives belonging to un-dominated or best categor(y)ies (Manzini and Mariotti (2012b)) or considering only top N eye-catching elements according to some exogenously given order or ranking (Salant and Rubinstein (2008)).

---

[6] This property is also implied by the appealing axiomatic rationalization theory of Cherepanov et al. (2013) (based on specific kinds of rationales), which supports its reasonability.

[7] This is not the case for unfiltered databases, where the result of their combinations is directly clear.

[8] For instance salience, familiarity, "roughly" identical features, heuristics, fairness, extremeness aversion, limited memory, reference effects etc..

Most of these approaches form consideration set in a way that can be interpreted as seeking for a reasons to select (based on Shafir et al. (1993) and Tversky (1972) and more related also in Lombardi (2009) and de Clippel and Eliaz (2012)).

Another approach to form a consideration set can be seen in a satisficing procedure (Simon (1959) and more related Tyson (2008, 2013), Papi (2012), Manzini et al. (2013a)). The resulting consideration set contains only these elements which surpass a (endogenously) given threshold level according to some criteria. This is close to our motivating similarity satisficing example in (2).

However, for our purpose the most interesting paper is the axiomatic rationalization theory of Cherepanov et al. (2013), since our assumed consideration property (of the filter) is a direct consequence of their normatively and descriptively appealing rationalization process. Modifying their justification procedure in order to cope with our other filter properties yields a corollary of our main representation result and allows for an interpretation in terms of a "rationalized" filtered belief.

The next section gives the general framework for the two stage filtered belief formation. In Section 3 we introduce and discuss the properties on the filtering process. Section 4 deals with the axioms on the belief formation induced by a filtered database. The main representation theorem and a sketch of the proof is presented in Section 5 and we derive as a corollary the similarity satisficing belief process. The last section relates the filtering process to the recently developed multi-criteria/stage decision procedures. In particular we exemplarily modify the choice model of Cherepanov et al. (2013) to a filter process in our terms. An interpretation of the resulting representation in terms of multi-similarities is given. Section 7 concludes. All proofs can be found in the appendix.

2   The model

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

2.1   Database framework

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

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

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

$D \circ E := (c_1, c_2, ..., c_L, c_1', c_2', ..., c_T')$.

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

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

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

Definition 2.1

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

(ii) The $\subseteq$-relation on the set of databases C^* is defined by $D \subseteq E \Leftrightarrow f_D(c)|D| \leq f_E(c)|E|$ for all $c \in C$. We will call such databases to be nested.

(iii) The $\cap$-relation on databases is given by $D \cap E = ((c^{\min\{f_D(c)|D|, f_E(c)|E|\}})_{c \in C})$

(iv) Two databases D and E are disjoint if for all $c \in C$: $c \in D$ if and only if $c \notin E$.

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

2.2   Filter

In the literature so far, a filter on C^* is usually defined as a set function $\Gamma : C^* \to C^*$, such that for all $D \in C^*$ $\emptyset \neq \Gamma(D) \subseteq D$. In this way the information D is filtered independent of the environment, additional information or prior knowledge. We call and summarize this additional source of information as the perspective from which an agent filters. Our definition of a filter allows different results of filtering, if the underlying perspective of evaluation changes. Formally, our filter depends also on another database E that acts as a perspective from which database D is filtered. A set function $\Gamma : X \times C^* \times C^* \to C^*$ in which for all $x \in X$ and $D \subseteq E \in C^*$, $\Gamma(x, D, E) := \Gamma^E(x, D) \subseteq D$ is called a filter on (sub-)database D induced by (perspective) database E and given a problem x. We use the notation $\Gamma^E(x, D)$ to highlight the roles of the two database. Basically the database D is screened from the perspective of having in mind the additional information E.

2.3   Belief

For a finite set S, $\Delta(S)$ denotes the simplex of probability vectors over S and for $n \in \mathbb{N}$ Δ^n denotes the simplex over the set $\{1, 2, ..., n\}$.

In the axiomatizations of BGSS, EG and Bleile (2014a), an agent will form a belief over outcomes $P(x, D) \in \Delta(R)$ for a certain problem characterized by $x \in X$ using her information captured in a database $D \in C^*$, i.e. $P : X \times C^* \to \Delta(R)$.

In the current approach a filtered belief is formed based only on parts of the information captured in D and that is filtered from the perspective of richer information $E \in C^*$, i.e. filtered belief $(P \circ \Gamma) : X \times C^* \times C^* \to \Delta(R)$ and $(P \circ \Gamma)(x, D, E) = (P \circ \Gamma^E)(x, D)$ for $D \subseteq E$ (with slight abuse of notation). In this sense the filtered belief $(P \circ \Gamma^E)(x, D)$ is induced by a nested pair of databases D and E and can be interpreted as the belief over

---

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

outcomes induced by database $D \in C^*$ seen through a filter that rests on perspective E (given problem $x \in X$). Hence, a filtered belief is a two stage process of filtering followed by a belief formation.

Technically, the filtered belief induced by the pair of nested databases D and E coincides with the BGSS belief based on an a priori already filtered database $\Gamma^E(D)$ [10]. However, as discussed in the introduction, a priori filtering would separate the filter procedure and the belief formation process that would exclude desirable applications based on endogenously derived similarity values, like the similar satisficing behavior (as discussed in our motivating example (equation 2)). Further, the nice axiomatization of BGSS based on relationships between filtered database and their induced beliefs would need to be defined on filtered databases as well, which merges both stages again. However, for our motivation and purpose the most intuitive (and desirable) filtered belief formation is based on a single database, i.e. the filtered belief $(P \circ \Gamma)(x, D, D)$ induced by database D.

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

3 Filter definition and properties

Instead of defining explicitly a procedure how to filter information we are more general and rather impose natural, well accepted and established properties in psychology and marketing. These properties are normatively and descriptively compelling in several situations and are indeed true for many heuristics people actually use in real life to screen their "information set". In particular many of the recently developed multistage decision models contain a wide variety of (endogenous and exogenous) filter procedures that satisfy our filter properties, which supports their relevance and generality (see Section 6). In this sense, we can also interpret our filter as a choice correspondence in which elements surviving the pre-choice process form a consideration set of acceptable and relevant information. This (filtered) consideration set represents the underlying basis for a filtered belief formation process.

3.1 Basic Properties of a filter

In the last section, we mentioned briefly our concept of a generalized filter. More precisely:

Definition 3.1 *Filter*
*A set function $\Gamma : C^* \times C^* \to C^*$ is called a **filter on D induced by (perspective) database** E on databases D, if for all $D, E \in C^*$ such that $D \subseteq E \in C^*$ it holds*
  (i) $\Gamma(E, E) =: \Gamma^E(E) =: \Gamma(E) \neq \emptyset$ and
  (ii) $\Gamma(D, E) =: \Gamma^E(D) = \Gamma(E) \cap D$.

Note that by definition $\Gamma^E(D) \subseteq D$ holds for all $D \subseteq E$, but it does not imply that $\Gamma^E(D) \neq \emptyset$.

As mentioned already, the traditional definition of a filter considers only one database, i.e. $\emptyset \neq \Gamma(D) \subseteq D$. Our definition is based on two nested databases $D \subseteq E$ for the filtering $\Gamma^E(D)$. Thereby, we want to capture the potential differences in filtering a database D

---

[10]Basically meaning that $(P \circ \Gamma^E)(D) = P(\Gamma^E(D))$.

depending on varying perspectives. Whether pieces of information receive attention might vary strongly with different perspectives of evaluations. However, our definition coincides with the usual traditional filter for any $D = E \in C^*$, i.e. for a situation in which no additional information than database D is available in agents' minds.

We will discuss the three properties the definition proposes, namely the filtering, its specific content and induced ordering.

There are many reasons, why an agent demonstrates partial attention to all available information. Cognitive limitations and constraints in parallel processing of complex and large information sets is often associated with an unconsciously allocation of attentional resources (Anderson (2005) pp.72-105). This kind of cognitive unawareness can lead to a formation of a consideration sets as well as an intentional heuristic elimination procedure in order to explicitly simplify the information set in a cost or effort efficient way, while still keeping the most relevant information. In addition, consideration sets are often constructed unrelated to the objective features of the elements in their information sets, e.g. subjective constraints, like motives, mood, rationales, value systems, biases, attitudes might affect acceptability, attractiveness or moral justification of an object and affects its assigned attention level. In sum, filtering a set of available information is a natural and plausible behavior under many circumstances and for many reasons.

The nested perspective-structure in our definition of the induced filter, i.e. $\Gamma^E(D) = \Gamma(E) \cap D$ for $D \subseteq E$, states that a database D is filtered from the perspective of a super-database E. In this specific form an agent already processed and filtered out the information in E (resulting in $\Gamma(E)$) that she deems relevant, wants to consider or is able to pay attention to. Since an agent actually needs to process and screen only information in a sub-database D, she will pay attention to all elements in D that already grabbed her attention and are still in her mind, i.e. $\Gamma(E) \cap D$. Basically D is not filtered independently but compared with the information that is attention grabbing in E. Put differently, D is filtered through E- colored glasses and not by inspecting elements in D in detail. However, it is important to stress that this definition applies only to situation in which both databases are processed "simultaneously". If databases are processed independently such an interwoven filtering (from the perspective of one) is not feasible or desirable. In particular, in the following we will introduce a (consideration) property which specifies how independently filtered nested databases are related.

Finally, we are concerned with the ordering of the resulting filtered database, which stems from the definition of the intersection and the ordering on C. Of course, it is restrictive to assume ad hoc a specific arbitrary ordering of a resulting filtered database, since amongst other things this may depend on orders of D or/and E. However, in course of the axiomatization of a belief formation, we would adopt the Invariance Axiom of BGSS, which states that only the content and not the ordering of cases is important for an induced belief. From that perspective any reordering of a resulting filtered database would lead to the same induced belief, which leaves the definition and its induced ordering rather harmless.

Moreover, the implied $\Gamma^E(D) = \Gamma^E(\pi(D))$ for all $D \subseteq E$ and any reordered database $\pi(D)$ will be further generalized by the following property.

Definition 3.2 *Filter order invariance*

*For all $L \in \mathbb{N}$ and $D \in C^L$ and any permutation π on $\{1,..,L\}$, let $D = (c_1,..,c_L)$ and $\pi(D) = (c_{\pi(1)},...,c_{\pi(L)})$. A filter Γ is **order invariant**, if it holds that $c \in \Gamma(D)$ if and only if $c \in \Gamma(\pi(D))$* [11].

Basically it states (in close relationship to BGSS Invariance Axiom for beliefs) that the order of the cases is immaterial for the resulting filtering of the cases, i.e. only the content of a database matters. Information should be attention grabbing per se and not due to its specific position in the database. From a first sight the property seems to be rather restrictive since agents appear to be able to consider all cases in the database "simultaneously" without any order biases, like first impressions and recency effect (see Rubinstein and Salant (2006)). However, if some of these effects are important, then they can be captured by a more elaborated description of the cases. For example, its description can include time or the order in which it was observed.

The filter order Invariance implies that in combination with the definition of intersections of databases

$$\Gamma^E(D) = \Gamma^{\pi'(E)}(\pi(D)) \text{ for all } D \subseteq E \text{ and any reordered databases } \pi(D) \subseteq \pi'(E), \quad (3)$$

It implies the identity of beliefs induced by all possible combinations of reordered databases $\pi(D)$ and $\pi'(E)$ that enter the filtering process.

A closely related property states and ensures quite naturally that if a case catches attention then all other cases of this type are attention grabbing as well.

Definition 3.3 *Equal treatment of information*
*A filter Γ on $C^* \times C^*$ **treats information equally**, if for all $c \in D$*

$$\Gamma^D(c) = c \text{ if and only if } \Gamma^D(c^T) = c^T \text{ for all } T \text{ such that } c^T \in D.$$

Basically, this means that all pieces of the same type are treated equally and either are attention grabbing or not.

Another content dependent property is similarly based on the view that attention or disregard is structural in a way that per se either a piece of information is eye-catching or attention grabbing or not. Namely, the filtering process is not affected by an (sufficiently large) amount of occurrence of a piece of information.

Definition 3.4 *Filter ignorance of repeated information*
*A filter Γ satisfies the **ignorance property** if for all $D \in C^*$ it holds:*
$\Gamma(D \circ c) \cap D = \Gamma(D)$ *for all $c \in D$* [12].

A repeated appearance of a case does not influence its perception in the sense that a case attracts attention if it is relevant or outstanding in itself and not because itself or another case appears in a specific amount. The assignment of attention to a case might be altered only if another new case appears or all appearances of a specific case are removed

---

[11]Note, that independent of the filter definition, the property would just enforce that the content is identical. Only in combination with our definition of a filter the ordering coincide, i.e. $\Gamma(D) = \Gamma(\pi(D))$ is implied.
[12]This is equivalent to: for all $D \subseteq E$ and all $c \in E$, $\Gamma^{E \circ c}(D) = \Gamma^E(D)$.

from the database (as induced by the consideration filter property defined below). In this sense, additional observation of already known and evaluated cases do not alter (already attached) attention levels.

The property finds support in Gul et al. (2012), which examines the probabilities of choices when alternatives are duplicated. They propose, that duplicate alternatives should be identified as observational identically and should be (in a specific sense) irrelevant for the likelihoods with which an observational identical alternative is chosen. This can be related to a "pay attention"- choice in which we restrict the probabilities to pay attention to zero or one. However, duplication of evidence might affect the composition of a filtered consideration set in the list-framework of Rubinstein and Salant (2006) and hence violate our property (and their partition independence property). Of course, the number of times an element appears might have an influence on attention, for instance in a procedure that pays attention only to the most frequent element.

3.2 Main structural properties of a filter

The two following structural properties characterize the impact of nested databases on their induced filtered consideration set and the cognitive ability of an agent to process and pay attention to at least a minimum amount of available information.

Consideration property

The consideration property specifies the relationship between the induced consideration sets of nested databases in a quiet naturally way. It states, that if an agent pays attention to a case in a database, then also her attention is drawn to this case in all its sub-databases. This follows the idea and evidence that elements in an information set compete for attention and need to outperform other pieces of information to be considered. If a case manages to attract attention or is salient enough in a database then it also gains attention in a sub-database, in which some of its rivals for attention are not anymore present. The classical example in marketing deals with the attention an agent assigns to products in a supermarket with a huge variety compared to a small neighborhood store. A specific jam catching your eye standing in front of a large supermarket shelf with fifty different jams will catch your attention also in the convenience store selling only five different sorts of jam.

Definition 3.5 *Consideration property*
A filter Γ *on* $C^* \times C^*$ *satisfies the* **consideration property***, if for all* $D \subseteq E \in C^*$ *and* $c \in D$ $c \in \Gamma(E)$ *implies* $c \in \Gamma(D)$.

From another point of view, if there is no reason (e.g. an (outstanding) piece of information) in the larger sample E that shadows case c, i.e. $c \in \Gamma(E)$, then it still cannot happen that case c is shadowed by any reason (e.g. any information) in the smaller sub-database.

Apart from the "competition for attention"-explanation the consideration property can be motivated by the finding that with increasing complexity and size of a set agents reduce the amount of alternatives they consider and also lower its intensity and seriousness [13]. If the complexity of a database is caused by a difficult detailed evaluation of alternatives

---

[13]See e.g. Hauser and Wernerfelt (1990) and Shugan (1980).

involving compromising and tradeoffs, an agent might find it much harder to find reasons why to "choose" to consider a case. As a consequence agents might stick to only superficial analysis, where only the very important, salient or extraordinary alternatives receive attention. With decreasing complexity of a set agents might return to a more detailed analysis that facilitates the selection of more alternatives to be worth or justified to be included into the consideration set.

Minimal attention span

The following property describes the cognitive ability of an agent to handle and process information. Our minimal attention property requires that an agent considers in full detail at least k-many ($k \geq 3$) available different pieces of information. More precisely, for all databases containing less than k many different information an agent takes into account all pieces of information. For more complex databases an agent is free to apply filtering techniques to process her information and neglects some pieces of information, as long as she pays attention to at least k-many different pieces of information.

Definition 3.6 *Minimal attention span*
*A filter is **minimal attentive** with $k \geq 3$, if for all $D \in C^*$ such that*
(i) for $div(D) = l \leq k$ it holds $div(\Gamma(D)) = l$ and
(ii) for $div(D) > k$ it holds $div(\Gamma(D)) \geq k$.

From this perspective, the property induces that agents are cognitively sophisticated enough to handle at least k-many different pieces of information completely and with full attention. Also, agents might want to gather a certain minimal amount of information to evaluate and act in an informed and confident way. Thus they take into account all available information, when only few (less than k different) information is around [14].

The property combines basically two components -minimum required amount of information and filtering in case of potential information overload - which are well supported by empirical findings. We state it in terms of a general minimal attention span k, but for our purpose $k = 3$ is sufficient and also meets empirical evidence. For example Gensch (1987) found that screening and filtering rules may be invoked for as few as four alternatives, but agents consider and rely on all information for less diverse information sets. In the marketing literature Jarvis and Wilcox (1977) examined the usual size of consideration sets and discovered that the average size of consideration sets is three to eight products, independent of the size of the initial information set [15].

So far a minimal attention property is usually not assumed in the (choice) literature in which usually no lower limit is given for the filtering stage unless a non-emptiness condition. In a special case of Lleras et al. (2010) agents have no limited attention problem for situations of binary alternative set, i.e. their agents are always able to pay attention to both alternatives [16]. However, their version does not cover any restrictions for more general situations, as specified in our property.

---

[14]One can also think about it as in Simon (1959, p.263) based on an aspiration or satisfaction level such that at least k cases are attracting interest or attention. If not, the level was too high and a reduction of the level leads to some search behavior of the agent to pay attention to more available alternatives.

[15]Miller' s insight (1956) that agents can process or remember at least seven case is also covered here.

[16]See also Masatlioglu et al. (2012), where $\Gamma(D) \geq 2$ allows the full revelation of preferences

The requirement of filtering a minimum amount of information constitutes a strong constraint to identify filters as choice correspondences. We will discuss this problematic issue in Section 6, where we interpret filtering in a choice theoretic perspective.

Definition 3.7 *Admissible Filter*
A filter Γ on $C^* \times C^*$ is called an admissible filter if it satisfies the invariance, equal treatment, ignorance, consideration and minimal attention span property.

4 The Axioms

In this section we introduce the axioms on the stage of belief formation.

4.1 Implied Filter Invariance Axiom

As mentioned in the section before, a natural axiom for a belief process that is already implied by the definition of intersections and the filter properties, is a version of an Invariance Axiom as in BGSS for filtered databases that reads:

Filtered Invariance Axiom (already implied)
For every $T \geq 1$, every database $D, E \in C^*$ with $D \subseteq E$. Let $\Gamma^E(D) = (c_1, .., c_T) \in C^T$ and for every permutation $\pi : \{1, ..., T\} \to \{1, ..., T\}$ and any filter $\Gamma^E(D)$, where $\pi(\Gamma^E(D)) = (c_{\pi(1)}, ..., c_{\pi(T)})$ the following holds:

$$(P \circ \Gamma^E)(D) = (P \circ \pi(\Gamma^E))(D))$$

Basically it says that an induced belief over outcomes depends only on the content of a filtered database and is insensitive to the sequence and order in which data arrives. However, by our definitions of a filter any filtered databases containing the same content have exactly the same specific ordering (according to the order on C), which makes the invariance property superfluous, since re-orderings do not occur after filtering (see equation (3)). In this sense the filtered Invariance Axiom is indirectly substituted by the definition of a filter and the filter ignorance property [17].

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

---

[17]We would have required the axiom directly, if our filtering process would have allowed for different orderings of databases consisting of the same content. Technically speaking, there is no difference between restricting the filtering process to specific orders or allowing for different orderings and requiring an Invariance Axiom for beliefs.

4.2 Filtered Concatenation Axiom

Let Γ be a filter. For all database $D, E, F \in C^*$ such that $D \circ E \subseteq F$ there exists $\lambda \in [0, 1]$ such that:

$$(P \circ \Gamma^F)(D \circ E)) = \lambda(P \circ \Gamma^F)(D) + (1 - \lambda)(P \circ \Gamma^F)(E)$$

where $\lambda = 0$ if and only if $\Gamma^F(D) = \emptyset$.

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

The filtered Concatenation Axiom says that a filtered belief induced by a concatenated database is a weighted average of the filtered beliefs induced by their respective combining databases. The axiom captures the idea that the belief based on the combination of two databases cannot lie outside the interval spanned by the beliefs induced by each combining database separately. Intuitively it can be interpreted in the following way (stated from an exclusion point of view): if the information in any database induces an agent's belief not to exclude an outcome r, then the outcome r cannot be excluded by the belief induced by the combination of all these databases [18]. Alternatively, if a certain conclusion is reached given two filtered databases, the same conclusion should be reached given their filtered union.

However, in order to sustain the normative appealing interpretation of averaging (filtered) beliefs, the filtered concatenation of two databases must coincide with the concatenation of these two filtered databases. This is achieved by employing the common perspective $F \supseteq D \circ E$ according to which all involved databases are filtered, i.e. Γ^F. Otherwise it would not be in general reasonable to require the existence of such an average of beliefs, since the elements surviving the filtering process for each single database might differ from the elements surviving the elimination of the database generated by the combination of the two [19]. In this situation, it would be implausible and unreasonable to determine a relationship between the induced filtered beliefs. The required structure ensures that a filtered belief induced by the concatenated database relies on information that is also employed in the filtered beliefs induced by the single concatenating databases and thus allows for an interwoven filtering and belief formation.

Moreover, another important reason for requiring this specific common perspective is based on the motivation that an agent should filter from the perspective of the richest information available, which is at least the concatenated database $D \circ E$, i.e. $F \supseteq D \circ E$. This is very reasonable and natural, since the agent processed already at least the information in the concatenated database due to the fact that she actually wants to form a (filtered) belief based on this filtered database. In this way she cannot remove (intentionally) some (gained or experienced) information for filtering the concatenating database from a less in-

---

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

[19]In an unfiltered concatenated database any information appear in either of their (unfiltered) concatenating databases and find weight in their induced beliefs. However for filtered databases $(P \circ \Gamma^{D \circ E})(D \circ E) = \lambda(P \circ \Gamma^D)(D) + (1 - \lambda)(P \circ \Gamma^E)(E)$ is meaningless, since the relationships between $\Gamma(D \circ E)$ and the parts $\Gamma(D)$ and $\Gamma(E)$ are unclear.

formative perspective [20]. In this way, the concatenated databases represent a quite natural choice as a "smallest" perspective from which the filtering process is initiated.

4.3 Collinearity Axiom

No three elements of $\{(P \circ \Gamma^c(c)\}_{c \in C}$ are collinear.

Technically speaking this axiom allows to derive a unique similarity function (in combination with the other axioms), but it has also some reasonable intuition. Roughly it states that the estimation based on a case is never equivalent to the combined estimations based on two other cases. Hence, a case is always informative in the sense that no combination of two other cases can deliver the same estimate and would make this case redundant.

5 Representation Theorem

Theorem 5.1
Let there be given a function $(P \circ \Gamma) : C^* \times C^* \to \Delta(R)$, where $P : C^* \to \Delta(R)$ and Γ be an admissible filter on $C^* \times C^*$. Let $(P \circ \Gamma) : C^* \times C^* \to \Delta(R)$ satisfy the Collinearity Axiom.
Then the following are equivalent:
(i) The function $(P \circ \Gamma)$ satisfies the filtered Concatenation Axiom
(ii) There exists for each $c \in C$ a unique $P^c \in \Delta(R)$, and a unique -up to multiplication by a strictly positive number- function $s : C \to \mathbb{R}_+$, such that for all $D \subseteq E \in C^*$ such that $\Gamma^E(D) \neq \emptyset$

$$(P \circ \Gamma^E)(D) = \frac{\sum_{c \in D} s(c) 1_{\Gamma(E)}(c) P^c}{\sum_{c \in D} s(c) 1_{\Gamma(E)}(c)} \tag{4}$$

5.0.1 Rough sketch of the proof

The necessity part is straightforward calculation. The sufficiency part follows the rough structure of the proof of BGSS and Bleile (2014a), but differs in the crucial arguments. The idea is to transform the framework from the space of databases to the space of frequency vectors that is structural more tractable, i.e. the filtered belief based on databases $(P \circ \Gamma^E)(D) = \frac{\sum_{c \in D} s(c) 1_{\Gamma(E)}(c) P^c}{\sum_{c \in D} s(c) 1_{\Gamma(E)}(c)}$ for $D \subseteq E \in C^*$ translates to frequency vectors $f_D \subseteq f_E$ by $(P \circ \Gamma^{f_E})(f_D) = \frac{\sum_{j \leq m} s_j \Gamma_j^{f_E}(f_D) P^j}{\sum_{j \leq m} s_j \Gamma_j^{f_E}(f_D)}$. In order to show that this is viable we exploit some properties of the filter and the Concatenation Axiom.
The essential part of the proof is to derive the similarity weights $(s_i)_{i \leq m}$. This will be shown inductively over $|C| = m$ and $div(f_E) \leq m$.

Step 1: Base case for the induction, i.e. $|C| = m = 3$ and $div(f_E) \leq 3$, w.l.o.g. $C = \{c_1, c_2, c_3\}$, i.e. aim to find s_1, s_2, s_3.
For pairs $(f_D, f_E) \in \Delta \times \Delta$ (i.e. such that $div(f_E) \leq 3$) the properties of the filter (in

---

[20]This is even impossible if the filtering occurs unconsciously, since the information entered already her mind by the nature of the task.

particular the minimal attention span) induce that the filtering stage disappears and the axioms coincide with BGSS. Thus, the same steps (using simplicial partitions) as in BGSS will show the above representation for all pairs (f_D, f_E) such that $f_D \subseteq f_E$ and $div(f_E) \leq 3$.

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

$$(P \circ \Gamma^{f_E})_s(f_D) = \frac{\sum_{j \leq m} s_j \Gamma_j^{f_E}(f_D) P^j}{\sum_{j \leq m} s_j \Gamma_j^{f_E}(f_D)} \text{ using the derived } s = (s_1, .., s_m)$$

The aim is to show $(P \circ \Gamma^{f_E})(f_D) = (P \circ \Gamma^{f_E})_s(f_D)$ for all $f_D \subseteq f_E \in \Delta$ via induction over m and using Step 1 ($m = 3$) as base case.
a) For all pairs (f_D, f_E) such that $div(f_E) = m$ and $div(f_D) < m$ the "idempotence" of the filter (i.e. $\Gamma(\Gamma^{f_E}(f_D)) = \Gamma^{f_E}(f_D)$) is used to exploit the induction assumption.
b) For $div(f_D) = div(f_E) = m$ we adopt (with different reasoning) the construction used in BGSS, i.e. $f_D = \alpha q^j + (1 - \alpha) f(j)$ (for some $\alpha \in (0, 1)$, where $f(j)$ denotes the point on $conv\{(q^l)_{l \in \{1,...,m\} \setminus j}\}$ on the line through f_D and f^j. The result of a) allows to apply the filtered Concatenation Axiom. The filter properties (minimal attention and consideration property) ensure that there exist at least three q^j such that $\Gamma^{f_D}(q^j) \neq 0^m$ and therefore there exist three P^j that do not lie on one line by the Collinearity Axiom. Thus, there are at least three different lines on which $(P \circ \Gamma^{f_D})(f_D)$ and $(P \circ \Gamma^{f_D})_s(f_D)$ lie and since their intersection is unique the beliefs $(P \circ \Gamma)$ and $(P \circ \Gamma)_s$ induced by the pair (f_D, f_D) have to coincide. The filter ignorance property concludes the proof for all (f_D, f_E).

5.0.2 Interpretation of Theorem

The main difference to the axiomatized representation of BGSS (i.e. (1)) lies obviously in the inclusion of the filtering process that captured by the indicator function in the representation (4). Thereby the agent only employs and needs to take into account the information that she really filtered to be most important, "relevant" or acceptable (according to some criteria, such that the admissible filter properties are met) [21]. Thus, the belief formation follows a two-stage procedure of filtering and subsequent belief formation. Another deviation from BGSS concerns the dependence of the axiomatization on pairs of databases and no single database as in BGSS. Such a structure is necessary and reasonable for our axiomatization. However for interpretational purpose and the motivation behind this work, the situation $E = D$ is most interesting. In this context, an agent is not concerned with additional information or priming of her mind by E, but evaluates the given or evoked database D. By interpreting a database D as her potentially available memory, an agent might recall or retrieve only some of her memorized experiences $\Gamma(D)$. Basically $\Gamma(D)$ can be seen as the result of a brainstorming or coming to mind process of those past experiences that she deems most appropriate, valuable, salient or wants to take into account.

---

[21]It allows also a conditional belief formation, i.e. $s(x, c) = 1_{\{x = x_c\}}(c)$, mentioned already in BGSS, but unfeasible in their setup.

5.1 Example: Similarity Satisficing

Initially Simon (1955, 1957) introduced satisficing behavior as an alternative approach to the classical rational choice theory. According to him, in most global models of rational choice all alternatives are evaluated before a choice is made, but " in actual human decision-making alternatives are often examined sequentially. We may, or may not, know the mechanism that determines the order of procedure. When alternatives are examined sequentially, we may regard the first satisfactory alternative that is evaluated as such as the one actually selected" (Simon (1955), p. 110). In general satisficing behavior is a relevant and often observed heuristic in reality (e.g. for experimental evidence see Caplin et al. (2011), Reutskaja et al. (2011)).

In our motivating example of a similarity satisficing procedure (2), the filtered belief was

$$(P \circ \Gamma^E)(D) = \frac{\sum_{c \in D} s(c) 1_{\{s(c) \geq s^*\}}(c) P^c}{\sum_{c \in D} s(c) 1_{\{s(c) \geq s^*\}}(c)}.$$

Its interpretation is especially appealing if a database is identified with recalled memory and assuming that those experiences are retrieved earlier that are most similar to the current problem. In this situation an agent will stop to contemplate after some time and starts to process the till-then recalled (most relevant) information.

Obviously, setting the threshold level to zero, i.e. $s^* = 0$, would result in the BGSS representation (1). This directly shows that s^* needs to be restricted in order to be meaningful embedded in our filtered belief formation-framework. In particular, our filtering process must satisfy the minimal attention and consideration property. The former property requires that we need to take into consideration the k ($k \geq 3$) most similar cases. This determines the threshold values s^*. Obviously, such a threshold needs to be database-dependent, i.e. $s^*(E) =: s^E$ for all $E \in C^*$. More precisely, define for all $E \in C^*$ $S_E := \{(s(c))_{c \in E}\}$ and denote by s^{E_j} the j-largest number $s(c)$ according to $\geq$ in S_E. Then we get directly that for all $E \in C^*$ the database-dependent threshold level s^E is given by $s^E = s^{E_k} 1_{\{div(E) \geq k\}}(E)$ for any cognitive ability $k \geq 3$. Such a definition of the threshold s^E implies the minimal attention property. In order to also satisfy the consideration property - i.e. $c \in \Gamma(E)$, then $c \in \Gamma(D)$ for all $D \subseteq E$ - we need to enforce $s^E \geq s^D$ for $D \subseteq E$. The resulting filter $\Gamma^E(D) = (c \in D \mid s(c) \geq s^E)$ satisfies the remaining properties directly.

Summarized, this yields the following Corollary.

Corollary 5.1

Let P be as in Theorem 5.1 and Γ a similarity satisficing filter $\Gamma^E(D) := (c \in D | s(c) \geq s^E)$ for $D \subseteq E \in C^*$ with database dependent similarity thresholds s^E as defined above. Then the equivalence in Theorem 5.1 holds with the specific representation

$$(P \circ \Gamma^E)(D) = \frac{\sum_{c \in D} s(c) 1_{\{s(c) \geq s^E\}}(c) P^c}{\sum_{c \in D} s(c) 1_{\{s(c) \geq s^E\}}(c)} \tag{5}$$

A database dependent threshold is even more in the spirit of Simon's satisficing behavior. Simons hypothesis is that most subjects search sequentially and stop search when an environmentally determined level of reservation utility (similarity in this context) has

been surpassed. Hence, for the specification of their reservation or satisficing level individuals take into account the environment, i.e. in our setup the given information set E. In addition, Simon proposed that the levels of reservation utility (similarity here) increase with set size and object complexity, i.e. for larger databases in our setup. Thus, both conditions on the threshold - database dependence and increasing in database complexity - are well-grounded in the satisficing literature.

A recent related paper that is concerned with the axiomatization of a two-stage threshold representation (Manzini et al. (2013a)) obtains various structures for the threshold values. Comparing it to Lleras et al. (2010) (or our admissible filter) they get as well $s^E > s^D$ for $D \subset E$. However the attention filter model of Masatlioglu et al. (2012) results in $s^E = s^D$ for any nested D and E, such that $\Gamma(E) \subseteq D$ and for the two stage salience model of Tyson (2013) even the converse inequality holds.

6 Related literature with consideration or elimination stage

6.1 Relationship to Multi-criteria or stage Decision Theory

As mentioned in the introduction the (implicit) formation of a consideration set is part of many recently developed multistage decision procedures in which a consideration set is constructed by eliminating several alternatives according to some criteria. The literature varies in the process of filtering. Some employ (sequences of) rational(es) to eliminate alternatives (Manzini and Mariotti (2007, 2012a), Apesteguia and Ballester (2013), Houy (2007, 2010), Houy and Tadenuma (2009)) or accept alternatives that can be justified by some of multiple criteria (Cherepanov et al. (2013) (CFS from now on), Gerasimou (2013)). Other procedures are based on undominated categor(y)ies (Manzini and Mariotti (2012b)) or specific frames (orders, lists, moods, fairness) which are unrelated to preferences (Salant and Rubinstein (2008)). Also mental constrains might induce consideration sets (Masatlioglu and Ok (2005)).

The aim of this section is to interpret and identify our filter in terms of a multi-criteria choice correspondence by adopting the above mentioned multistage elimination procedures. A problematic issue in merging both concepts lies in the fact that choice models usually are intended to identify a single chosen alternative by choosing in the final step the "best" alternative within the remaining consideration set. However, for a filtering process and implied corresponding consideration sets singletons are not desirable. For this reason we do not identify and compare the entire choice process with a filtering process, but we are mainly interested in the filtering and elimination stages and not on the final choice stage. On the other hand we can stick to these choice models, if we replace the criteria applied in the final choice step (often binary asymmetric relation) by appropriate satisficing criteria (as discussed above) such that we end up with a set of acceptable alternatives. Roughly speaking, we discuss and relate the models in a more approximative and intuitive style, being aware of the difficulties and basic differences.

6.2 Filter as choice correspondence in multistage procedures

In order to identify a filter Γ as a choice correspondence we need to discuss our filter properties in a multistage decision theoretic framework.

The **filter definition** $\emptyset \neq \Gamma(D) \subseteq D$ is plausible for a choice correspondence since active, non-empty choices need to be made from D. The second part of the definition - $\Gamma^E(D) = \Gamma(E) \cap D$- can be interpreted as a usual consistency condition or as choice from E given a (budget) constraint D.

Since usual decision theoretic frameworks deal with sets of alternatives in which orderings and repetitions are immaterial, the **properties of invariance, equal treatment** and **ignorance of additional identical information** are directly satisfied for a choice correspondence.

Our **minimal attention span** property can be interpreted as a restriction to multiple choices (i.e. correspondences) such that a minimum of k available cases need to be chosen. Of course this requirement differs from common decision theoretic frameworks in which no restrictions on the quantity of chosen elements is enforced (unless non-emptiness). Thus, it will be crucial to implement this property into an adopted choice correspondence. A possible approach will be proposed and discussed in Section 6.2. In particular we exemplarily adopt CFS's rationalization model, but the taken approach can be applied to other models as well (discussed below). Another approach to guarantee for a minimum amount of choices can be imposed by enforcing an appropriate satisficing strategy at some stage of the choice process.

However, for the moment we want to focus on the **Consideration property**. In the choice literature it is known as Sen's property α or akin as Contraction Axiom. Many modified versions of the weak Axiom of revealed preferences (WARP) satisfy the consideration property. Manzini and Mariotti (2007) introduce a **Weak WARP** that is also satisfied by CFS (2013), Lleras et al. (2010) and Lombardi (2009) that states that for all $c, c' \subset D \subset E$ and $c = \Gamma((c, c')) = \Gamma(E)$, then $c' \notin \Gamma(D)$. This coincides with a **ReWARP** of Gerasimou (2013). Lleras et al. (2010) introduce a **Limited Consideration WARP** that states that for any $c \in D \cap E$ it holds that $c \in \Gamma(E)$ if (i) $\Gamma(E) \in D$ and (ii) $c \in \Gamma(A)$ for some $A \supseteq E$.

However, there are also some refinements of WARP in our context that do not satisfy the consideration property, e.g. Gerasimou (2013)'s **DeWARP** is neither satisfied in generality nor **Moody WARP** of Manzini, Mariotti (2013b).

For the remaining section our interest lies on the elimination procedures suggested in the multistage decision models and their link to the consideration property. We translate the choice theoretic approaches directly into our database framework and will not state the original versions of their models.

6.2.1 Models without an explicit procedure to form consideration sets

In Lleras et al. (2010) the formation of the consideration set is directly characterized by the consideration property, as in our approach.

Masatlioglu et al. (2012) axiomatizes choice based on a consideration set generated by

an attention filter, which is characterized by

$$\Gamma(D) = \Gamma(D \backslash c) \text{ for all } c \in D \backslash \Gamma(D)$$

Basically, such a filter selects only those cases in a database D that she is aware of and effectively pays attention to. Hence, if an agent is not aware of $c \in D \backslash \Gamma(D)$, then removing c from the database D should not affect the set of cases an agent would pay attention to. In contrast, our consideration property does not rely on such an unawareness component and in general both properties (attention and consideration) are independent (see Lleras, et al. (2010) for an example demonstrating the differences).

6.2.2 Models with explicit formation procedures

Given a binary relation R on C, we denote by $U(D, R) := (c \in D \mid \, / \exists \tilde{c} \text{ with } \tilde{c}Rc)$ [22] the cases in D that are undominated in D and by $Dom(D, R) := (\tilde{c} \in D \mid \exists c \in D \backslash \tilde{c} \text{ such that } cR\tilde{c})$ the set of cases in D that are dominated by a case in D.

Sequential elimination Procedures

The following approaches adopt the same rough idea of "short-listing" in which multi-criteria are checked sequentially and only those alternatives survive until the final consideration stage that meet some or all criteria. The elimination is based mainly on un-dominance inspections regarding the specific criterion.

Manzini and Mariotti (2007) axiomatize choice following a rational shortlisting behavior. The final choice is made according to criterion R_2 within the alternatives in the consideration set that are undominated and survived the elimination according to R_1 (asymmetric and transitive binary relation)

$$\Gamma(D) = (c \in D \mid c \in U(R_2, U(R_1, D))).$$

Houy (2007, 2010) introduces as well another mechanism to form a consideration set that is based on a composition of some binary relations R_i, in which agents sequentially check for a certain pattern of (un)dominance according to ordered criteria

$$\Gamma(D) = (c \in D \mid \text{ for all } \tilde{c}, \ cR_1\tilde{c} \text{ or } (\neg(\tilde{c}R_1c) \text{ and } cR_2\tilde{c}) \text{ or } (\neg(\tilde{c}R_2c) \lor cR_3\tilde{c}...)).$$

In a similar spirit Horan (2013) summarizes many of these two stage models in which a consideration stage is formed according to undominance based on a asymmetric relation and then a second (asymmetric) relation is used for the choice.
Obviously, all such short-listing procedures (and extended to more criteria $(R_i)_i$) do satisfy the consideration property.

A sequential elimination procedure of a different kind is discussed in Manzini and Mariotti (2012b). Their consideration set is formed according to a (asymmetric, possibly

---

[22]For our purpose it might make more sense to think rather in terms of satisficing than maximal elements, e.g. in the sense of $U(R_i, D) = (c \in D \mid cR_ic_i^*)$ for some threshold case c_i^*.

incomplete) relation P on subsets that are interpreted as categories. The un-dominated categories survive the elimination phase.

$$\Gamma(D) = (S \subset D \mid \not\exists \tilde{S} \subset D \text{ such that } \tilde{S}PS).$$

In order to capture the consideration property, specific requirements on the categorization structure are necessary. Bleile (2014b) implements two potential versions of categorizations on databases into a belief formation process.

Satisficing procedures

The following branch of literature adopts the satisficing idea of Simon (1955, 1957).

Tyson (2008) axiomatizes a satisficing procedure based on a considerations set that contains only those elements that exceed some database dependent threshold level Θ according to some numerical representation of a criterion f

$$\Gamma(D) = (c \in D \mid f(c) \geq \Theta(D)).$$

Such a filter satisfies the consideration property for an appropriate definition of the database dependent threshold value Θ (as in our similarity satisficing example in Section 5.1).

Tyson (2013) and Manzini et al. (2013a) modify and generalize this procedure to salience measures and general relations.

Papi (2012) proposes an axiomatic characterization of the satisficing heuristic under various informational structures in which the order of inspecting alternatives are either full, partially or not observable. Especially the case of unobserved sequences can be interpreted within a framework of choice correspondences by assuming that for all possible orders the satisficing elements enter the consideration set.

Frame related elimination procedures

Salant and Rubinstein (2008) model a general approach in which the in principle available choice set is restricted by subjective and psychological constraints, rationales or biases. They call such additional characteristics that are not directly covered by the objective description of the alternatives as frames

$$\Gamma(D) = (c \in D \mid \Gamma(D, f) = c \text{ for some frame } f).$$

Obviously, for a general unspecified frame our filter properties are not directly satisfied.

Reason Based Choice procedures

In general, the above mentioned procedures to form a consideration set can be interpreted under the premise of having and/or seeking a reason to accept or eliminate alternatives. This general idea follows the stream of literature on reason based choice initiated by Shafir et al. (1993) or even Tversky (1972). That is, elements in the consideration set are those

that can be (internally) justified most easily (according to some reasons). Thereby as above, an (un)dominance structure (according to one, some or all criteria) serves as convincing reason for choosing the specific element. A link between reason based choice and the consideration property can be established by the insight that in a smaller set it might be easier to find a reason to choose some alternative, whereas in a larger set it also might be easier to find a reason to reject.

Lombardi (2009) relies on the concept of reason based choice in the sense of finding the best and most easily justifiable alternatives by possessing the "most convincing " dominance structure. It constructs a consideration filter by employing the same criterion for the screening as well as for the final evaluation, but in different ways [23]. The rational (acyclical binary relation) used as a criterion in the first stage is used to construct another criterion for the second stage.

$$\Gamma(D) = (c \in U(D) \mid \nexists \, \tilde{c} \in U(D) \text{ such that } Dom(\tilde{c}, D, R) \supset Dom(c, D, R)),$$

where $Dom(c, D, R) := (\tilde{c} \in D \text{ such that } cR\tilde{c})$. Such a procedure does not satisfies the consideration property in general.

Gerasimou (2013) also relies on a procedure based on a single (acyclical or asymmetric) relation R. The consideration set contains elements that are justified by the fact that they are un-dominated, but at least one alternative is worse [24]

$$\Gamma(D, R) = (c \in D \mid \nexists \, \tilde{c} \in D \text{ such that } \tilde{c}Rc \text{ and } \exists c' \in D \text{ s. th. } cRc').$$

Such a procedure does not satisfy in general the consideration property.

Similarly, in the vein of seeking reasons to justify the selection, De Clippel and Eliaz (2012) employ a pro-cons bargaining procedure based on linear orders P_1, P_2. The agent forms the consideration set via (internal) compromising between P_1 and P_2 by trying to receive as many as possible dominated alternatives for both rationales.

$$\Gamma(D) = (c \in D \mid argmax_c min_i |Dom(D, P_i)|)$$

The consideration property needs not to hold in general for this procedure.

However, the most interesting and elegantly fitting model for our approach is CFS's (2013) Rationalization Theory that we want to merge with our approach exemplarily.

6.2.3 Rationalization Theory and related psychological filter

CFS model a consideration filter explicitly as a rationalization procedure. For a set of general binary relations $R = (R_1, ..., R_n)$ a case in database D is rationalized if $cR_i\tilde{c}$ for all $\tilde{c} \in D$ for some $i \leq n$. This psychological filter contains those alternatives that are

---

[23]In general, the psychological literature (e.g. Andrews and Srinivasan (1995), Roberts and Lattin (1997)) states that the criteria influencing consideration and the final evaluation stage may differ as well as (partially) overlap. Overlapping criteria, however, play different roles at both stages.

[24]Note, there is a difference to undomination

justifiable by at least one criterium, rational, reason, story, etc.

$$\Gamma^R(D) = (c \in D \,|\, \exists i \leq n \text{ such that } cR_i\tilde{c} \text{ for all } \tilde{c} \in D).$$

This procedure is very interesting, since such a psychological filter satisfies the consideration property directly and hence the rationalization procedure can be seen as a generator of any filter satisfying the consideration property. However, for our purpose we need to take care of our additional minimal attention property. Roughly speaking, we want to find some reasonable rationales $S = (S_1, ..., S_N)$ that deliver an admissible filter via a rationalization procedure similar to CFS.

For a binary relations S and for $D \in C^*$, we define the following recursive (maximal) domination sets for an attention level $k \geq 3$

$$Max(D,S) =: Max^1(D,S) := (c \in D \,|\, cS\tilde{c} \text{ for all } \tilde{c} \in D)$$
$$\text{for } n > 1: \quad Max^n(D,S) := Max(D \setminus \cup_{i \leq n-1} Max^i(D,S), S) \quad \text{and}$$
$$Max^*(D,S) := Max^d(D,R) \text{ for } d := argmin_n\{div(Max^n(D,S)) \geq k\}.$$

Note, that $Max^*(D,S)$ can be empty, e.g. for $div(D) < k$.

Definition 6.1
a) Let $S = \{S_1, ..., S_N\}$ be a be a set of binary relations on C and $D \in C^*$. Then we call a case c minimal k-attentive rationalizable (MAR) by S in D if and only if
$c \in Max^*(D,S_1) \lor c \in \circ_{i=2}^N Max(D,S_i)$.
The set of all MAR cases in database D by rationales S is denoted by $MAR(D,S)$.

b) A filter Γ is called a minimal k-attentive psychological filter based on a set of binary relations $S = \{S_1, .., S_N\}$, i.e. $\Gamma = MAR(D,S)$, if it holds
(i) for any $D \in C^*$ such that $div(D) \leq k$: $\Gamma(D) = D$ and
(ii) for any $D \in C^*$ such that $div(D) \geq k$ and $Max^*(D,S_1) \neq \emptyset$: $\Gamma(D) = MAR(D,S)$.

Basically the modification of a CFS filter serves the reason to capture the required minimal attention property by enforcing ad hoc that always the k-best cases according to a most important, seminal, distinguishing, leading rational (criterium, reason, story) S_1 are consider for sure. In addition, if they differ from the best cases according to the other criteria, also all cases which are rationalizable by these other rationales survive the elimination procedure. A plausible way to justify such a formation process would emphasize the extraordinary role of criterion S_1. An agent is focussing on the (at least) k-best undominated or most salient alternatives for the most important criterion S_1 and only the best alternatives according to the minor, rather marginal or negligible criteria are worth to consider. For instance, an agent buying a car would choose according to different criteria, like speed, mileage, gas consumption, etc. Her major criteria might be gas consumption and hence includes the k best cars regarding economy into her consideration set, whereas she only takes the fastest car and that with lowest mileage into account, since they are outstanding or salient within the minor criteria.

For the definition and the underlying recursive domination we have in mind $k = 3$. For larger minimal attentions k one can generalize this approach to any specific structure

of ranking criteria. For instance for $k = 4$, one might assume to consider for sure the two best alternatives according to rational S_1, and additional the two best remaining alternatives regarding to story S_2. In this sense a minimal attentive psychological filter can be generalized in arbitrary ways for specific attention levels.

The non-emptiness requirement of a MAR-filter, i.e. $Max^*(D, S_1) \neq \emptyset$, is for example satisfied if the binary relation S_1 is complete. But also for an incomplete "benchmark" or satisficing relation S_1 [25] the non-emptiness is satisfied if c^* is chosen such that there exist at least k cases $c \in D$ such that cS_1c^*

Consequently, we can state the following corollary

Corollary 6.1

Let Γ be an MAR-filter on $C^* \times C^*$ based on a set of binary relations $S = \{S_1, ..., S_N\}$. Then Γ is an admissible filter and the equivalence in the Theorem 5.1 holds for all $D \subseteq E \in C^*$, such that $\Gamma^E(D) \neq \emptyset$ with the specific representation

$$(P \circ \Gamma^E)(D) = \frac{\sum_{c \in D} s(c) 1_{MAR^*(E,S)}(c) P^c}{\sum_{c \in D} s(c) 1_{MAR^*(E,S)}(c)}$$

In general, most of the above mentioned multistage procedures can be adopted to satisfy the minimal attention property by replacing an usual dominance structure by our defined Max^* structure for some rational at some stage of their elimination procedures. In this way, these modified multistage choice processes (that satisfy the consideration property) can be interpreted as an admissible filter and incorporated into our filtered belief formation.

An appealing and intuitive example for rationales S_i in S is to interpret them as rationales that are related to componentwise similarities on the characteristics space $X = X_1 \times ... \times X_N$, i.e. $s_i(c) \in \mathbb{R}_+$ for all $c \in C$ ($i \leq N$). One can understand the endogenously derived similarity value s as an (complex) aggregation f (f: $\mathbb{R}^N \to \mathbb{R}$) of these componentwise similarities- i.e. $s(c) = f(s_1(c), ..., s_N(c))$. An underlying motive for choosing such a form relies on the fact that agents tend to evaluate the dimensions lexicographically instead of aggregating multi-evaluation criteria (Tversky et al. (1988), Dulleck et al. (2011)), i.e. $(s_i)_{i \leq N}$ versus $s = f(s_1, .., s_N)$. In addition, research shows that in the filtering stage agents use noncompensatory heuristics for a rough screening based on simple criteria, e.g. a satisficing behavior like comparing s_i with thresholds s_i^*. But for the final evaluation stage a compensatory, more detailed multi-component and compromise-based procedure is taken, as in our aggregated similarity measure $s = f(s_1, .., s_N)$ [26].

Two candidates for a reasonable and plausible, but ad hoc, definition of binary relations S_i are based on comparisons of the componentwise similarities s_i in the following way.

Definition 6.2

Let there be functions $s_i : C \to \mathbb{R}$ for all $i \leq N$.

(i) "Componentwise similarity": For all $i \leq N$, a transitive and complete binary relation

---

[25] I.e. for a benchmark c^* $c^* S_1 c'$ if and only if $cS_1c^* S_1c'$ or $(cS_1c^*$ and $c'S_1c^*)$ for $c, c', c^* \in D$, see also Definition 6.2 later.

[26] See e.g. Bettman (1979), Gensch (1987), Hauser and Wernerfelt (1990), Roberts and Lattin (1991, 1997), Andrews and Srinivasan (1995) and for a model Lombardi (2009).

$\tilde{S}_i$ is defined on $C \times C$ by $c\tilde{S}_i c'$ if and only if $s_i(c) \geq s_i(c')$

(ii) "Benchmark exceeding componentwise similarity": For all $i \leq N$ the asymmetric, transitive and possibly incomplete binary relation S_i^* is defined by $cS_i^*\tilde{c}$ if and only if $(s_i(c) \geq s_i^* > s_i(\tilde{c}) \quad \vee \quad s_i(c) \geq s_i(\tilde{c}) \geq s_i^*)$ for componentwise threshold values $(s_i^*)_i$.

Obviously, the binary relation defined in (i) can be used to define a MAR-filter. The relations in (ii) can be applicable for a MAR-filter if for any $D \in C^*$ s_1^* is chosen according to $s_1^* \leq s_1^{D_k} 1_{\{div(D) \geq k\}}(D)$ (as defined in Section 5.1).

7   Conclusion

This paper examines how beliefs are formed by agents that are constraint or not willing to pay attention to all potentially available pieces of information. It is well known in the psychology and marketing literature that humans do not take into account all available information due to many different reasons. Based on this insight we axiomatize a two stage belief formation procedure in which agents employ only these pieces of information that "survived" a first step of (un(intentional)) filtering (or screening). The filter is required to satisfy natural, reasonable and well known properties. The axioms on the belief level are closely related to the axioms introduced in BGSS and modified in a way to capture the link to filtering information and their consequences for induced beliefs. The resulting filtered belief is a weighted sum of estimates induced by past observed information that are attention grabbing. Thus only pieces of information that attracted the attention and consideration of an agent are taken into account. The weights are determined by the similarities of the observed cases with the problem under consideration.

The axiomatized filtered belief formation generalizes the axiomatizations of BGSS, EG and Bleile (2014a) in which all available pieces of information are necessarily taken into account which prevents unintentional forgetting or unawareness as well as intentional application of a heuristic screening techniques, that often drive human judgment. Hence, a filtered belief formation offers a cognitively less demanding and maybe more realistic behavioral procedure to form beliefs based on data.

An intuitive and natural application of a filtered belief formation are models of satisficing behavior regarding the relevance or appropriateness of information for the current problem. Moreover, it captures also a conditional belief formation process that only takes into account identical problems in the past and neglects all not perfectly similar observations - which cannot be covered by BGSS, EG and Bleile (2014a).

In particular interesting is that filtering (and elimination) of information (or alternatives) emerged very recently as a research topic in the decision theory literature. These multi-stage/criteria models incorporate as well a first step of filtering before engaging into the final choice step. Many of these models can be easily translated and embedded into the filtering stage of our belief formation process such that our filter stage can be interpreted as a choice correspondence in terms of decision theory.

A Proof of Theorem 5.1, necessity part (ii) $\Rightarrow$ (i)

Let $(P \circ \Gamma^E)(D) = \frac{\sum_{c \in D} s(c) 1_{\Gamma(E)}(c) P^c}{\sum_{c \in D} s(c) 1_{\Gamma(E)}(c)}$ for all $D \subseteq E \in C^*$, such that $\Gamma^E(D) \neq \emptyset$.
Let $D = D_1 \circ D_2$ and $D \subseteq E \in C^*$.

$$
\begin{aligned}
(P \circ \Gamma^E)(D) &= \frac{\sum_{c \in D} s(c) 1_{\Gamma(E)}(c) P^c}{\sum_{c \in D} s(c) 1_{\Gamma(E)}(c)} \\
&= \frac{1}{\sum_{c \in D} s(c) 1_{\Gamma(E)}(c)} \Big(\sum_{c \in D_1} s(c) 1_{\Gamma(E)}(c) P^c + \sum_{c \in D_2} s(c) 1_{\Gamma(E)}(c) P^c \Big) \\
&= \sum_{i=1,2} \frac{\sum_{c \in D_i} s(c) 1_{\Gamma(E)}(c)}{\sum_{c \in D} s(c) 1_{\Gamma(E)}(c)} \Big(\frac{\sum_{c \in D_i} s(c) 1_{\Gamma(E)}(c) P^c}{\sum_{c \in D_i} s(c) 1_{\Gamma(E)}(c)} \Big) \\
&= \lambda (P \circ \Gamma^E)(D_1) + (1 - \lambda)(P \circ \Gamma^E(D_2)
\end{aligned}
$$

Thus, the the filtered Concatenation Axiom is satisfied. $\qquad \square$.

B Proof of Theorem 5.1, sufficiency part (i) $\Rightarrow$ (ii)

B.1 Important Observations

In the proof, we treat the situation in which the level of minimal attention k is set equal to three, i.e. $k = 3$. This simplifies notational effort and is sufficient to follow the main steps of the proof that analogously work for any $k \geq 3$.

The following Lemma states useful and crucial properties for the proof.

Lemma B.1
Let Γ be an admissible filter, then the following holds:
(i) For all $c \in D \in C^*$, $\Gamma^D(c^T) \in \{\emptyset, c^T\}$ for all T such that $c^T \in D$
(ii) For all $D \subseteq E \in C^*$ such that $div(E) \leq 3$: $D \subseteq \Gamma^E(D) \subseteq D = ((c^{f_D(C)|D|})_{c \in C})$
(iii) For all $D \subseteq E \in C^*$ we have some kind of idempotence, i.e.

$$
\Gamma(\Gamma^E(D)) = \Gamma^E(D)
$$

Proof:
(i) By definition $\Gamma^D(c) \subseteq c$, hence the equal treatment property delivers directly the desired result.

(ii) By definition and the equal treatment property, we have for all $D \subseteq E \in C^*$ that $\Gamma^E(D) = \circ_{c \in C}(\Gamma^E(c))^{f_D(c)|D|}$
If $div(E) = k \leq 3$, then by minimal attention property $div(\Gamma(E)) = k$ and thus by the equal treatment property $\Gamma(E) = \circ_{c \in C} c^{f_E(c)|E|}$. Since $D \subseteq E$, we get directly $\Gamma^E(D) = \circ_{c \in C} c^{f_D(c)|D|}$.

(iii) By definition of a filter, we have $\Gamma(\Gamma^E(D)) \subseteq \Gamma^E(D)$. The consideration property implies
$\Gamma(\Gamma^E(D)) = \Gamma^{\Gamma^E(D)}(\Gamma^E(D)) \supseteq \Gamma^E(\Gamma^E(D)) = \Gamma^E(E) \cap \Gamma^E(D) \supseteq \Gamma^E(D)$
and hence the claim holds true. $\qquad \square$

Basically (ii) just says that no filtering of D takes place in this situation (only a reordering takes place, which is "immaterial" for the induced belief).

B.2 Translating the database framework to frequencies

An essential step in the proof is to identify databases with their frequency vectors. The space of frequency vectors is more tractable and enables us to adopt the structure of BGSS's proof (and use the procedure of of Bleile (2014a)). However, the proof here requires some additional features, since in addition filters are involved, which alters the crucial steps in the inductive proof.

B.2.1 General Definitions for a Frequency Framework

We need to introduce some definitions regarding the frequency framework.

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

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

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

Definition B.1
(i) O^m denotes the null-vector on $\mathbb{R}^m$.
(ii) For all $j \in \{1, 2, ..., m\}$ denote by q^j the j-th unit vector in $\mathbb{R}^m$, i.e. the frequency vector representing a database containing only case $c_j \in C$, i.e. $q^j = (0, ..., 0, \underbrace{1}_{j-th}, 0, ..., 0)^t$

(iii) For all $d \in \Delta$ its diversity is given by $div(d) := |\{i \leq m \,|\, d_i > 0\}|$

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

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

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

$$d \cap e := ((\frac{\min\{d_i T, e_i L\}}{\sum_{i \leq m} \min\{d_i T, e_i L\}})_{i \leq m}) \ ^{27}$$

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

B.2.2 Why is a transformation viable?

Roughly, we want to show that for a filtered belief formation we can identify databases $D \subseteq E \in C^*$ in the filtering process by frequencies $f_D \subseteq f_E$ such that $\Gamma^{f_E}(f_D)$ corresponds to $\Gamma^E(D)$. For this purpose, we exploit the properties of an admissible filter and the axioms on filtered belief formation in the following way.

(i) The filter ignorance property for Γ implies directly $\Gamma^{E^L}(D) = \Gamma^E(D)$ for $D \subseteq E$ and for all $L \in \mathbb{N}$, i.e. $(P \circ \Gamma^E)(D) = (P \circ \Gamma^{E^L})(D)$.
(ii) The filtered Concatenation Axiom implies (by $D^Z = D \circ ... \circ D$) $(P \circ \Gamma^F)(D) = (P \circ \Gamma^F)(D^Z)$ for an appropriate F such that $D^Z \subseteq F$ holds for $Z \in \mathbb{N}$.

---

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

However, since for all $Z \in \mathbb{N}$ there exists a $L \in \mathbb{N}$ such that $D^Z \subseteq F^L$, observations (i) yields that for $D \subseteq E$

$$(P \circ \Gamma^E)(D) = (P \circ \Gamma^{E^L})(D^Z) \text{ for all } Z \in \mathbb{N} \text{ and all sufficiently large } L \in \mathbb{N} \qquad (6)$$

Further, by the definition of a filter and its order invariance property (and hence its implied belief invariance property) the order of the cases do not matter, which enables us to represent all involved databases by their frequency vector on C (which are independent of lengths) and their corresponding lengths.

However, by the above observations, within the filtered belief formation, lengths of databases become irrelevant in the sense of equation (6). In particular, since each $D^Z \in C^*$ (for any $Z \in \mathbb{N}$) is represented by the same frequency vector f_D, the "sufficiently large"-condition looses its bite. Thus, we can identify the filtered belief formation process on databases by frequency vectors.

Filtered belief induced by frequencies

Definition B.3
The filtered beliefs $(P \circ \Gamma) : C^* \times C^* \to \Delta(R)$ based on databases $D \subseteq E$ translates to corresponding beliefs based on frequency vectors $d \subseteq e$ in the following way:
$(P \circ \Gamma) : \Delta \times \Delta \to \Delta(R)$ such that $\Gamma(f_D, f_E) := \Gamma(D, E)$ and $(P \circ \Gamma)(f_D, f_E) := (P \circ \Gamma)(D, E)$.

Basically, the weakening of the condition $D \subseteq E$ to $f_D \subseteq f_E$ runs through the implicit or intuitive interpretation of $\Gamma^{f_E}(f_D)$ in a way such that there exists an appropriate replication Z of database E (i.e. $f_E = f_{E^Z}$) such that $E^Z \supseteq D$ is matched and since by the ignorance property $\Gamma^{E^Z}(D) = \Gamma^E(D)$ the nestedness condition can be relaxed.
Thus an application of filter properties and belief axioms show the viability of the transformation from databases to the frequency framework.

B.2.3   Filter definition in frequency terms

Definition B.4
For $d \subseteq e \in \Delta$, a function $\Gamma: \Delta \times \Delta \to \Delta$ is called an e-induced filter on d, if
(i) $\Gamma(e,e) := \Gamma^e(e) := \Gamma(e) \in \Delta$ and (ii) $\Gamma^e(d) := \Gamma(e) \cap d$ hold.

Definition B.5
*(i) **Consideration property:***
A filter Γ on $\Delta \times \Delta$ satisfies the consideration property if for $d \subseteq e \in \Delta$: $\Gamma(e) \cap d \subseteq \Gamma(d)$

*(ii) **Minimal attention span:***
A filter Γ satisfies the minimal attention of $k \geq 3$, if for all $d \in \Delta$:
(a) If $div(d) = l \leq k$, then $div(\Gamma(d)) = l$
(b) If $div(d) > k$, then $div(\Gamma(d)) \geq k$.

*(iii) **Filter ignorance of repeated information:***
Let $d, e = (\frac{p_1}{p}, \frac{p_2}{p}, ..., \frac{p_m}{p}), f = (\frac{p_1}{p+1}, \frac{p_2}{p+1}, .., \frac{p_i+1}{p+1}, ..., \frac{p_m}{p+1}) \in \Delta$, where $p := \sum_{j \leq m} p_j \in \mathbb{N}$.
Let $d \subseteq e \subseteq f$. A filter Γ satisfies the ignorance property, if $\Gamma^e(d) = \Gamma^f(d)$.

The **equal treatment** property is directly satisfied by the definition of a filter in frequency terms.

Definition B.6
A filter Γ on $\Delta \times \Delta \to \Delta$ satisfying the consideration, equal treatment, minimal attention span and ignorance property is called admissible.

Analogously to Lemma B.1, we get in frequency terms:

Lemma B.2 *(Lemma B.1 in frequency terms)*
Let Γ be an admissible filter on $\Delta \times \Delta$.
(i) For all $q^j \subseteq d \in \Delta$, we have $\Gamma^d(q^j) \in \{\emptyset, q^j\}$.
(ii) For all $d \subseteq e \in \Delta$ such that $div(e) \leq 3$ $\Gamma^e(d) = d$ holds.
(iii) For all $d \subseteq e \in \Delta$: $\Gamma(\Gamma^e(d)) = \Gamma^e(d)$

B.2.4 Axioms in frequency terms

Filtered Concatenation Axiom
Let Γ be a filter on $\Delta \times \Delta$. For all $d \in \Delta_T$ and $e \in \Delta_L$ for any $T, L \in \mathbb{N}$, there exists $\lambda \in [0, 1]$, such that for $g \supseteq f := \frac{T}{T+L}d + \frac{L}{T+L}e$

$$(P \circ \Gamma^g)(f) = \lambda(P \circ \Gamma^g)(d) + (1 - \lambda)(P \circ \Gamma^g)(e),$$

where $\lambda = 0$ if and only if $\Gamma^g(d) = O^m$.

Collinearity Axiom
No three of $\{(P \circ \Gamma^{q^j})(q^j)\}_{j \leq m}$ are collinear.

B.2.5 Sufficiency part of Theorem 5.1 in frequency terms

Proposition B.1
Let there be given a function $(P \circ \Gamma) : \Delta \to \Delta(R)$, where $P : \Delta \to \Delta(R)$ and Γ an admissible filter on $\Delta \times \Delta$. Let a filtered belief $(P \circ \Gamma) : \Delta \times \Delta \to \Delta(R)$ satisfies the filtered Concatenation and Collinearity Axiom.
Then, there exist unique probability vectors $(P^j)_{j \leq m} \in \Delta(R)$, and unique -up to multiplication by a strictly positive number- strictly positive numbers $(s_j)_{j \leq m} \in \mathbb{R}$ such that for all $q \subseteq f \in \Delta$ such that $\Gamma^f(q) \neq (0, ..., 0)$

$$(P \circ \Gamma^f)(q) = \frac{\sum_{j \leq m} s_j \Gamma_j^f(q) P^j}{\sum_{j \leq m} s_j \Gamma_j^f(q)} \tag{7}$$

where $\Gamma_j^f(q)$ denotes the frequency of case c_j in $\Gamma^f(q)$.

B.3 Proof: Sufficiency part of Theorem 5.1 in frequency terms

We have by Lemma B.2 (ii) directly that $\Gamma^f(q^j) = q^j$ for all $f \supseteq q^j$ such that $div(f) \leq 3$ and hence we need to choose

$$P^j = (P \circ \Gamma^f)(q^j) \tag{8}$$

The aim of the inductive proof over m with $|C| = m$ or $div(f) \leq m$ is to find the similarity values $s_1, ..., s_m$.
Step 1: $|C| = m = 3$, w.l.o.g. $C = \{c_1, c_2, c_3\}$, **thus** $\Delta = \Delta(q^1, q^2, q^3)$
Step 1.1: Defining similarity weights
We define $q^* := \frac{1}{3}(q^1 + q^2 + q^3)$ and for $f \supseteq q^*$ Lemma B.2 (ii) yields

$$(P \circ \Gamma^f)(q^*) = \frac{\sum_{j \leq 3} s_j \Gamma_j^f(q^*) P^j}{\sum_{j \leq 3} s_j \Gamma_j^f(q^*)} = \frac{\sum_{j \leq 3} s_j \frac{1}{3} P^j}{\sum_{j \leq 3} s_j \frac{1}{3}} = \frac{\sum_{j \leq 3} s_j P^j}{\sum_{j \leq 3} s_j}$$

According to the filtered Concatenation Axiom there exist $\lambda \in int(\Delta^3)$ (by minimal attention, i.e. $\Gamma^f(q^j) = q^j$) such that

$$(P \circ \Gamma^f)(q^*) = \sum_{j \leq 3} \lambda_j (P \circ \Gamma^f)(q^j) = \sum_{j \leq 3} \lambda_j P^j,$$

where the last equality follows from (8).

By equating both representations we can derive the corresponding similarity weights s_1, s_2, s_3 uniquely up to multiplication by a strictly positive number and define for all $q \subseteq f \in \Delta$

$$(P \circ \Gamma^f)_s(q) = \frac{\sum_{j \leq m} s_j \Gamma^f_j(q) P^j}{\sum_{j \leq m} s_j \Gamma^f_j(q)}$$

The aim is now to show that for all $(q, f) \in \Delta \times \Delta$ such that $q \subseteq f$

$$(P \circ \Gamma^f)_s(q) = (P \circ \Gamma^f)(q). \tag{9}$$

All such (q,f) are collected in $E := \{(q, f) \in \Delta \times \Delta \mid (P \circ \Gamma^f)_s(q) = (P \circ \Gamma^f)(q)\}$, where obviously by definition

$$\begin{aligned} (q^j, f) &\in E \qquad \text{for all } j \leq 3 \text{ and } f \in \Delta \text{ such that } q^j \subseteq f \\ \text{and } (q^*, f) &\in E \qquad \text{for all } f \in \Delta \text{ such that } q^* \subseteq f. \end{aligned} \tag{10}$$

Step 1.2: All simplicial points (with appropriate perspective) satisfy equation (9)

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). The main tool of the proof is the following observation, which will be recursively applied in an appropriate manner in the proof.

Lemma B.3

Let $a, b, c, d, e \in \Delta$, where $e = (a, b) \cap (c, d)$ and for all $f \in \{a, b, c, d\}$ let $div(f) \leq 3$ and $(f, f) \in E$. If $((P \circ \Gamma^f)(f))_{f \in \{a,b,c,d\}}$ are not collinear, then $(e, g) \in E$ for $g \in \Delta$ such that $div(g) \leq 3$ and $e, f \subseteq g$ for all $f \in \{a, b, c, d\}$.

Proof:

W.l.o.g. let e be between a and b on the line through a and b. By the filtered Concatenation Axiom we get

$$(P \circ \Gamma^e)(e) \in ((P \circ \Gamma^e)(a), (P \circ \Gamma^e)(b)) \text{ and } (P \circ \Gamma^e)_s(e) \in ((P \circ \Gamma^e)_s(a), (P \circ \Gamma^e)_s(b)),$$

since $(P \circ \Gamma)_s$ and $(P \circ \Gamma)$ satisfy the axiom. For $f \in \{a, b\}$ such that $div(f) \leq 3$, Lemma B.2 (ii) generalizes it to

$$(P \circ \Gamma^g)(e) \in ((P \circ \Gamma^g)(a), (P \circ \Gamma^g)(b)) \text{ and } (P \circ \Gamma^g)_s(e) \in ((P \circ \Gamma^g)_s(a), (P \circ \Gamma^g)_s(b))$$

for all $g \in \Delta$ such that $div(g) \leq 3$ and $e, f \subseteq g$.

Analogously, we get a similar result for the segment (c, d).

By Lemma B.2 we know that $\Gamma^g(f) = \Gamma^f(f)$ for all $f \in \{a, b, c, d\}$ and since by assumption $(f, f) \in E$, i.e. $(P \circ \Gamma^f)_s(f) = (P \circ \Gamma^f)(f)$, we directly get $(P \circ \Gamma^g)_s(f) = (P \circ \Gamma^g)(f)$ and

$$(P \circ \Gamma^g)(e), (P \circ \Gamma^g)_s(e) \in ((P \circ \Gamma^g)(a), (P \circ \Gamma^g)(b)) \cap ((P \circ \Gamma^g)(c), (P \circ \Gamma^g)(d)).$$

Since the intersection of the two line is unique due to the Collinearity Axiom, we get the desired result, i.e. $(P \circ \Gamma^g)(e) = (P \circ \Gamma^g)_s(e)$ and $(e, g) \in E$. $\qquad \square$.

By Lemma B.2 we know that $\Gamma^f(q) = q$ for any $q \subseteq f$ such that $div(f) \leq 3$, hence we need to show equality (9) not for all appropriate pairs (q, f), but only for any q and some appropriate f such that $q \subseteq f$. Then it will hold for all f such that $q \subseteq f$.

In the following, we will partition the simplex Δ into so called simplicial triangles re-

cursively, as illustrated in the Figure 1 below.

Definition of Simplicial Triangles:

The 0-th simplicial partition consist of vertices $q_0^j \in \Delta$, which are exactly the unit vectors q^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. $q_1^1 := (\frac{1}{2}q^1 + \frac{1}{2}q^2)$, $q_1^2 := (\frac{1}{2}q^2 + \frac{1}{2}q^3)$ and $q_1^3 := (\frac{1}{2}q^3 + \frac{1}{2}q^1)$. 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.
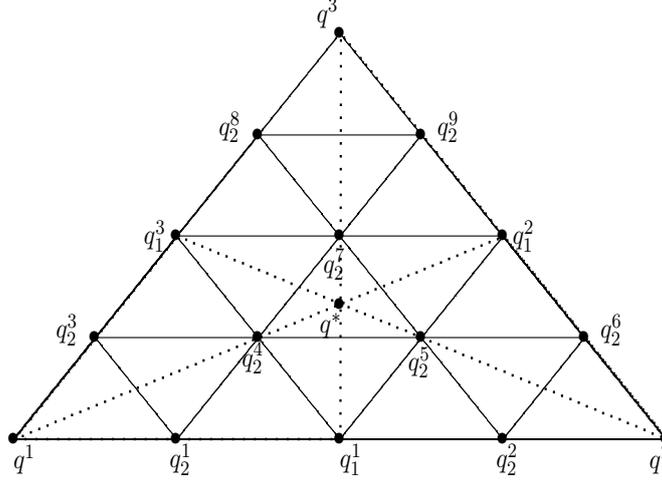
Figure 1: 1st and 2nd Simplicial partitions

We want to show that all simplicial points satisfy equation (9), i.e. are in E, by induction over the l-th simplicial partitions. Step 1.1. showed the claim for $l = 0$ (equation (10)). We proceed to the points in the
First simplicial partition:
Since
$$q_1^1 = (q^1, q^2) \cap (q^3, q^*),$$
and we already know that for all $f \in \{q^1, q^2, q^3, q^*\}$ $(f, f) \in E$ we can apply Lemma B.3 if the collinearity condition holds. However, since $(P \circ \Gamma^{q^i})(q^i) = P^i$ for $i = 1, 2, 3$ and $(P \circ \Gamma^{q^*})(q^*) \in int(conv(\{P^1, P^2, P^3\}))$ the Collinearity Axiom directly induces the non-collinearity condition. Hence, by Lemma B.3 we get that $(q_1^1, f) \in E$ for all f such that $q_1^1 \subseteq f$.

With the same reasoning, we get $(q_1^2, f), (q_1^3, f) \in E$ where $f \supseteq q_1^2$ (respectively q_1^3). Thus all pairs (q_1^i, f) consisting of a simplicial points of the first simplicial partition and all f such that $q_1^j \subseteq f$ are included in E.

For the **second simplicial partition** we distinguish between inner simplicial points and points on the boundary of the simplex Δ, i.e. between two of the corners q^j. Figure 2 demonstrates the intuition.

(a) The **first step** involves the inner simplicial points $q_2^4, q_2^5, q_2^7 \in int(conv(\{q^1, q^2, q^3\}))$.

Since
$$q_2^4 \in (q_1^1, q_1^3) \cap (q^1, q_1^2)$$
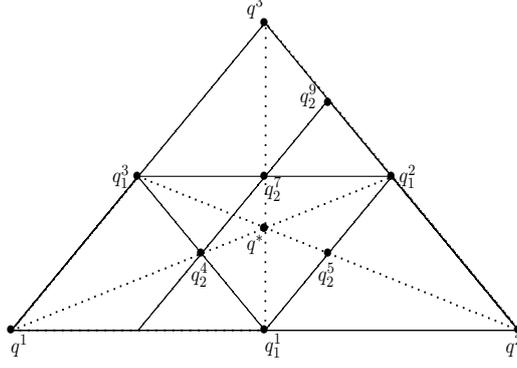
Figure 2: Step from 1st to 2nd Simplicial partition

and for all $f \in \{q_1^1, q_1^3, q^1, q_1^2\}$ $(f, f) \in E$ by Step 1.1 and Step 1.2 for the first simplicial partition, we need to check the collinearity condition to apply Lemma B.3. However, the condition is met since the induced beliefs for (q_1^j, q_1^j) with $j = 1, 2, 3$ are in $int(conv(\{P^1, P^j\}))$ that cannot lie on one line since $(P^j)_{j \leq 3}$ are not collinear by the Collinearity Axiom. Consequently by Lemma B.3 we get that $(q_2^4, f) \in E$ for all $f \supseteq q_2^4$.

Analogously, we can get that all simplicial points (combined with appropriate perspective f) of the 2nd partition in the interior of Δ, i.e. q_2^5, q_2^7 with appropriate super-frequencies f are in E.

(b) In the **second step** we will deal with and focus on the simplicial points on the boundary of Δ (e.g. representative q_2^9, see Figure 2).

We have that $q_2^9 \in (q^3, q^2) \cap (q_2^4, q_2^7)$. All frequencies f involved in the intersection are shown (Step 1.1. and Step 1.2 (a) for second partition) to be contained in E, in the sense of $(f, f) \in E$. Again, the non-collinearity is fulfilled since $(P^2$ and $P^3)$ and induced beliefs in $int(conv(\{P^1, P^2, P^3\}))$ are involved and $(P^2, P^3) \notin int(conv(\{P^1, P^2, P^3\}))$ since $(P^j)_{j \leq 3}$ are not collinear. Thus, Lemma B.3 delivers $(q_2^9, f) \in E$ for all $f \supseteq q_2^3$.

The same procedure with analogous and adjusted arguments yield that all simplicial points on the boundary of the 2nd simplicial partition combined with appropriate super-frequencies (perspectives) are also included in E.

The same kind of algorithm works for all simplicial points of any l-th simplicial partitions, i.e. obviously each $q \in rim(conv(\{q^1, q^2, q^3\}))$ is for some l captured. For $q \in int(conv(\{q^1, q^2, q^3\}))$ one can approximate q via a series of simplicial points (q_l^1, q_l^2, q_l^3) such that $q \in int(conv(\{q_l^1, q_l^2, q_l^3\}_l))$ for all l. In detail, the completion for all permissible $(q, f) \in \Delta \times \Delta$ can be shown almost similarly as in Step 1.3 Bleile (2014a) (or differently in BGSS Step 1.2 in their proof) and hence we refer to these papers for the entire procedure. This concludes the proof for the case $|C| = 3$ or (q, f) with $q \subseteq f$ such that $div(f) \leq 3$.

Now we need to show the claim for $|C| = m > 3$ or $div(f) \leq m$.
Step 2: $|C| = m > 3$
Step 2.1: Defining the similarity weights
Using the considerations from Step 1 above for $\{j, k, l\}$ (i.e. $q^*_{\{j,k,l\}} := \sum_{i \in \{j,k,l\}} q^i$, $q^i \in \Delta$ and $q \subseteq f \in \Delta(q^j, q^k, q^l)$) we can derive the similarity weights $(s_i^{\{j,k,l\}})_{i \in \{j,k,l\}}$. Further, for all (q,f) such that $q \subseteq f \in conv(\{q^j, q^k, q^l\})$ the following representation holds

$$(P \circ \Gamma^f)^{\{j,k,l\}}(q) = \frac{\sum_{i=j,k,l} s_i^{\{j,k,l\}} \Gamma_i^f(q) P^{\{j,k,l\}}(q^i)}{\sum_{i=j,k,l} s_i^{\{j,k,l\} \Gamma_i^f(q)}}$$

32

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

Similar to BGSS or Bleile (2014a), we can show that the similarity values $s_i^{\{j,k,l\}}$ are independent of the choice of j,k and l for all $i \in \{j,k,l\}$, since filtering is not present in the arguments. Thus we can define for all $q \subseteq f \in \Delta$

$$(P \circ \Gamma^f)_s(q) := \frac{\sum_{i \leq m} s_i \Gamma_i^f(q) P^i}{\sum_{i \leq m} s_i \Gamma_i^f(q)}$$

The aim is to show that for all $(q,f) \in \Delta \times \Delta$ such that $q \subseteq f$ $(P \circ \Gamma^f)(q) = (P \circ \Gamma^f)(q)$.

Step 2.2: Completion to all $(q,f) \in \Delta \times \Delta$

By Step 1 we know that the claim $(P \circ \Gamma^f)_s(q) = (P \circ \Gamma^f)(q)$ is true for all (q,f) such that $div(f) \leq 3$. We take this as the base case of our induction.

For the induction assumption, we have that $(P \circ \Gamma^f)_s(q) = (P \circ \Gamma^f)(q)$ for all $(q,f) \in \Delta \times \Delta$ with $q \subseteq f$ and $div(f) \leq k - 1$.

The induction step considers $q,f \in \Delta$ with $q \subseteq f$ and $div(f) \leq k$:

We can restrict the analysis to f such that $div(f) = k$, since for all other $f \in \Delta$ the claim is true by the induction assumption.

We split the proof into two parts. First for which $div(q) \leq k - 1$ and then for $div(q) = k$.

First Situation: Consider $q \subset f$, i.e. $div(q) \leq k - 1$.

By Lemma B.2 (iii), we have $\Gamma(\Gamma^f(q)) = \Gamma^{\Gamma^f(q)}(\Gamma^f(q)) = \Gamma^f(q)$ and hence directly

$$(P \circ \Gamma^f)(q) = (P \circ \Gamma^{\Gamma^f(q)})(\Gamma^f(q)), \tag{11}$$

Since $\Gamma^f(q) \subseteq q$ by definition of a filter and hence $div(\Gamma^f(q)) \leq k-1$ the induction assumption applies to the RHS of equation (11), i.e. $(P \circ \Gamma^{\Gamma^f(q)})(\Gamma^f(q)) = (P \circ \Gamma^{\Gamma^f(q)})_s(\Gamma^f(q))$ which is again identical to $(P \circ \Gamma^f)_s(q)$ and hence the desired result $(P \circ \Gamma^f)_s(q) = (P \circ \Gamma^f)(q)$ is implied directly.

Second Situation: Consider $q \subseteq f$ with $div(q) = k$

A similar construction as in BGSS, but with different reasoning, yields the result. Let $q = \sum_{l \in K} \alpha_l q^l$ with $\alpha_l > 0$ and $K \subseteq \{1, ..., m\}$ such that $|K| = k$.

Define the frequency vector $q(l)$ to be the vector in $conv(\{(q^j)_{j \in K \setminus l}\})$ such that q lies on the line $(q(l), q^l)$.

By the minimal attention span property $div(\Gamma^q(q)) = div(\circ_{\{l:q_l>0\}} \Gamma^q(q^l)) \geq 3$, i.e. there exist at least three l's (e.g. $l = i,j,k$) such that $\Gamma^q(q^l) = q^l \neq \emptyset$ and hence $(P \circ \Gamma^q)(q^l) = P^l = (P \circ \Gamma^q)_s(q^l)$.

Further, for these $l \in \{i,j,k\}$ we get $(P \circ \Gamma^q)_s(q(l)) = (P \circ \Gamma^q)(q(l))$ by the result of the first situation, since $div(q(l)) \leq k - 1$.

Hence we have that $(P \circ \Gamma^q)_s(q), (P \circ \Gamma^q)(q) \in (P^l, (P \circ \Gamma^q)(q(l))) =: L(l)$, for those three $l = i,j,k$. Since no three P^j are collinear, there are at least two distinct lines L(l), i.e. $L(l) \neq L(n)$ for at least two distinct $l,n \in \{i,j,k\}$. Since $(P \circ \Gamma^q)_s(q), (P \circ \Gamma^q)(q)$ are both on these distinct lines and these lines need to intersect uniquely, we have $(P \circ \Gamma^q)_s(q) = (P \circ \Gamma^q)(q)$. By the ignorance property $(P \circ \Gamma^q)(q) = (P \circ \Gamma^f)(q)$ for all f with $div(f) = k$, which completes the proof. $\square$.

References

[1] Anderson. J.R. (2005). Cognitive Psychology and Its Implications. *Worth, New York.*

[2] Andrews, R.L., Srinivasan, T.C. (1995). Studying Consideration Effects in Empirical Choice Models Using Scanner Panel Data. *Journal of Marketing Research, 32-1, 30-41.*

[3] Apesteguia, J., Ballester, M. (2013). Choice by Sequential Procedures. *Games and Economic Behavior, 77, 90-99.*

[4] Bettman, J.R. (1979). Issues in Research on Consumer Choice. *Advances in Consumer Research, 6, 214-217.*

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

[6] Bleile, J. (2014a). Cautious Belief Formation. *Working Paper, Center for Mathematical Eonomics, Bielefeld 507*

[7] Bleile, J. (2014b). Categorization based Belief Formation. *Unpublished paper.*

[8] Broadbent, D. E. (1958). Perception and Communication. *Pergamon Press, New York.*

[9] Caplin, A., Dean, M., and Martin, D. (2011). Search and Satisficing. *American Economic Review, 101-7, 2899-2922.*

[10] Cherepanov, V., Feddersen, T., Sandroni, A. (2013). Rationalization. *Theoretical Economics, 8, 775-800.*

[11] Chiang, J., Chib, S., Narasimhan, C. (1998). Markov chain monte carlo and models of consideration set and parameter heterogeneity. *Journal of Econometrics, 89-1-2, 223-248.*

[12] de Clippel, G., Eliaz, K. (2012). Reason-based choice: A bargaining rationale for the attraction and compromise effects. *Theoretical Economics, 7, 125-162.*

[13] Dekel, E., Lipman, B. (2012). Costly Self-Control and Random Self-Indulgence. *Econometrica, 80-3, 1271-1302.*

[14] Dulleck, U., Hackl, F., Weiss, B. Winter-Ebmer, R. (2011). Buying online: An Analysis of Shopbot Visitors. *German Economic Review, 12-4, 395-408.*

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

[16] Eliaz, K., Spiegler, R. (2011a). Consideration sets and competitive marketing. *Review of Economic Studies, 78, 235–262.*

[17] Eliaz, K., Spiegler, R. (2011b). On the Strategic Use of Attention Grabbers. *Theoretical Economics, 6, 127-155.*

[18] Gensch, D. H. (1987). A Two Stage Disaggregate Attribute Choice Model. *Marketing Science, 6, 223-31.*

[19] Gerasimou, G. (2013). A Behavioural Model of Choice in the Presence of Decision Conflict. *School of Economics and Finance Discussion Paper No. 1302, University of St Andrews.*

[20] Gigerenzer, G., Todd, P. and the ABC Research Group (1999). Simple Heuristics That Make Us Smart. *Oxford University Press, New York.*

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

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

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

[24] 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.*

[25] Gul, F. and W. Pesendorfer (2001). Temptation and Self-Control. *Econometrica, 69, 1403-1435.*

[26] Gul, F., Natenzon, P., Pesendorfer, W. (2012). Random Choice as Behavioral Optimization. *Working Paper.*

[27] Hauser, J. R., Wernerfelt, B. (1990). An evaluation cost model of evoked sets. *Journal of Consumer Research, 16, 383-408.*

[28] Hauser, J. R. (2013). Consideration Set Heuristics. *Forthcoming, Journal of Business Research.*

[29] Horan, S. (2013). A Simple Model of Biased Choice. *Working paper, forthcoming: Journal of Economic Theory.*

[30] Houy, N. (2007). Rationality and order-dependent sequential rationality. *Theory and Decision, 62, 119-134.*

[31] Houy, N., Tadenuma, K. (2009). Lexicographic compositions of multiple criteria for decision making. *Journal of Economic Theory, 144, 1770-1782.*

[32] Houy, N. (2010). A characterization of prudent choices. *Social Choice and Welfare, 34-2, 181-192.*

[33] Jarvis, L., Wilcox, J. (1977). A model for predictive measurements of advertising effectiveness. *Industrial Marketing Management, 6, 9-14.*

[34] Lleras, J.S., Masatlioglu, Y., Nakajima, D., Ozbay, E.Y. (2010) and (2012). When more is less: Limited consideration. *Unpublished paper.*

[35] Lombardi, M. (2009). Reason-Based Choice Correspondences. *Mathematical Social Sciences, 57, 58-66.*

[36] Manzini, P., Mariotti, M. (2007). Sequentially Rationalizable Choice. *American Economic Review, 97, 1824-1839.*

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

[38] Manzini, P., Mariotti, M. (2012b). Choice by lexicographic semiorders. *Theoretical Economics, 7, 1-23.*

[39] Manzini, P., Tyson, C. J., Mariotti, M. (2013a). Two-stage threshold representations *Theoretical Economics, 8-3, ?.*

[40] Manzini, P., Mariotti, M. (2013b). Moody Choice. *Working Paper.*

[41] Masatlioglu, Y., Ok, E. A. (2005). Rational Choice with Status Quo Bias. *Journal of Economic Theory, 121, 1-29.*

[42] Masatlioglu, Y., Nakajima, D., Ozbay, E.Y. (2012). Revealed attention. *American Economic Review, 102, 2183-2205.*

[43] Miller, G. A. (1956). The magic number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review, 63, 81-97.*

[44] Nedungadi, P. (1990). Recall and Consumer Consideration Sets: Influencing Choice Without Altering Brand Evaluations. *Journal of Consumer Research, 17-3, 263-76.*

[45] Papi, M. (2012). Satisficing choice procedures. *Journal of Economic Behavior and Organization, 84-1, 451-462.*

[46] Pessemier, E. A. (1978). Stochastic properties of changing preferences. *American Economic Review, 68-2, 380-85.*

[47] Reutskaja, E., Nagel, R., Camerer, C. F., Rangel, A. (2011). Search Dynamics in Consumer Choice Under Time Pressure: An Eye-Tracking Study. *American Economic Review, 101-2, 900-926.*

[48] Roberts, J. H., Lattin, J. (1991). Development and Te,sting of a Model of Consideration Set Composition *Journal of Marketing Research, 28, 429- 440.*

[49] Roberts, J. H., Lattin, J. (1997). Consideration: Review of research and prospects for future insights. *Journal of Marketing Research, 34, 406- 410.*

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

[51] Salant, Y., Rubinstein, A. (2008). (A,f): Choice with Frames *Review of Economic Studies, 75, 1287-1296.*

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

[53] Schwartz, B. (2004). The Paradox of Choice: Why More is Less. *Harper Perennial, New York.*

[54] Shafir, E., Simonson, I., Tversky, A. (1993). Reason-based choice. *Cognition, 49, 11–36.*

[55] Shugan, S. M. (1980). The cost of thinking. *Journal of Consumer Research, 27-2, 99-111.*

[56] Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics, 69-1, 99-118.*

[57] Simon, H. A. (1957). Models of Man:Social and Rational: Mathematical Essays on Rational Human Behavior in a social setting. *New York: John Wiley and Sons, Inc.*

[58] Simon, H. A. (1959). Theories of Decision-Making in Economics and Behavioral Science. *The American Economic Review, 49-3. 253-283.*

[59] Stigler, G.J. (1961). The economics of information. *Journal of Political Economy, 69-3, 213-225.*

[60] Tversky, A. (1972). Elimination by Aspects: A Theory of Choice. *Psychological Review 79, 281-299.*

[61] Tversky, A., Sattath, S., Slovic, P. (1988). Contingent weighting in judgment and choice. *Psychological Review, 95, 371-384.*

[62] Tyson, C.J. (2008). Cognitive constraints, contraction consistency, and the satisficing criterion. *Journal of Economic Theory, 127-1, 51-70.*

[63] Tyson, C. J. (2013). Satisficing Behavior with a Secondary Criterion. *Working Paper.*

[64] Wright, P., Barbour, F. (1977). Phased Decision Strategies: Sequels to an Initial Screening. *In Multiple Criteria decision Making, 91109. Amsterdam: North-Holland Publishing Co.*