On Deciding When to Decide*

Itzhak Gilboa[†]and Fan Wang [‡]

May 2017

Abstract

We consider a decision maker who follows a status quo without reconsidering her implicit decision at every period. Only as a result of certain events will she ask herself whether she would like to change her choice. We ask when this mode of decision making is compatible with optimality. We state conditions on the set of databases that would make the decision maker take an explicit decision, which are equivalent to the following representation: the decision maker entertains a set of theories, of which one is that her current choice is the best; she is inert as long as that theory beats any alternative theory according to a maximum likelihood criterion.

1 Introduction

Some decisions are taken, and some just happen. For example, assume that Mary wakes up in the morning and goes to work. She could have decided to quit her job on this day and start a job search. An outside observer who follows the revealed preference paradigm would view Mary as if she made a decision of not quitting, even if the thought did not cross her mind that particular morning. However, there are circumstances under which this

*We thank for comments on earlier versions of this paper. Gilboa gratefully acknowledges ISF Grant 704/15.

[†]HEC, Paris-Saclay, and Tel-Aviv University. tzachigilboa@gmail.com

[‡]HEC, Paris-Saclay. fan.wang1@hec.edu

choice would be conscious and explicit. For example, Mary may receive some feedback on her job performance, positive or negative, which may make her wonder whether she won't be better off in a different job or with a different employer. She might have heard of a friend who quit a similar job for a new one, or who changed his career altogether. Similarly, we may consider a person who, implicitly, makes a daily decision not to emigrate. Most days the decision would not be thought of as such, but certain events in the personal, professional, or political domains might prod him to consider the possibility of emigration consciously.

In this paper we consider the implicit decision to make an explicit decision. We do not model the latter, but only the former. That is, we do not study the acts the decision maker would consider once she "wakes up" and makes a conscious decision. We only consider the binary decision of "make a conscious decision and check if the status quo seems to be optimal" vs. "remain inert and go about my business". While we tend to think of this decision as implicit and unconscious, it need not be irrational. A person who faces such a decision might engage in explicit, fully rational decision process and end up with the conclusion that the status quo is indeed her optimal choice. Moreover, along the lines of "rational inattention" (Sims, 2003), in many situations it might even be "more rational" to stick to the status quo without reconsidering it than to incur the cognitive and emotional costs involved in conscious decisions. Indeed, when considering the choice of a career or a country to live in, there is typically little that changes from day to day to warrant a full-fledged, conscious and explicit decision process.

The question we ask is, when can one rationalize the implicit decision not to make an explicit one? Specifically, we consider a decision maker who (implicitly) entertains a set of theories, of which one, t_0, is interpreted as "my current choice is optimal", and others suggest that various alternatives are better than the status quo. Each theory is represented by a statistical model, assigning probabilities to "cases". Given a database of cases, the decision maker's choice to stick to the status quo is rationalizable if t_0 is a maximizer of the likelihood function.

Cases in our model are abstract entities that may correspond to various pieces of information. In the job search example, cases may involve information about the decision maker, such as positive or negative reviews; about others, such as the fact that another person changed her job; or about no one in particular, such as the news that there is higher demand for certain skills in certain markets. In the emigration decision example, a case might be the emigration of another, a hate crime in one's country, a statement by a politician, and so forth. Thus, cases may or may not be accounts of similar decision problems (faced by the self or by others), and if they are, they may or may not specify outcomes and payoffs.

As in Gilboa and Schmeidler (2003, 2010) we consider databases that allow for possible repetitions of cases that are considered to be identical as far as the decision maker is concerned. A database is therefore modeled as a "counter vector", attaching an integer $I(c) \geq 0$ to each "case type" c, signifying the number of times cases of this type have been encountered. The set of all conceivable databases is divided into two – the set in which the decision maker "wakes up" to make a conscious decision, $\mathcal{D}$, and the set in which she sticks with the status quo without pondering her choice, $\mathcal{S}$. We state conditions on $\mathcal{S}$ and $\mathcal{D}$, which are equivalent to the following representation: the decision maker entertains a set of "theories", T, each of which is a statistical model that assigns probabilities to cases. That is, for $t \in T$, $t(c)$ is the probability theory t assigns to a case of type c. A database I is in $\mathcal{S}$ if an only if

$$\prod_c [t_0(c)]^{I(c)} \geq \prod_c [t(c)]^{I(c)}$$

for all $t \in T$.

Thus, the decision maker is "inert", namely, does *not* make a conscious decision, precisely for the set of databases given which the theory t_0 is at least as likely as all the others. We also generalize the representation to allow for a priori preferences for some theories over others, which are combined with the likelihood function as in Akaike Information Criterion (Akaike, 1974).

The rest of this paper is organized as follows. The next sub-section is devoted to a discussion of related literature. Section 2 presents the model and the main result. Some extensions are presented in Section 3, while Section 4 contains a discussion.

1.1 Related Literature

Simon (1957) and March and Simon (1958) suggested the model of "satisficing", according to which a decision maker doesn't engage in conscious optimization, but rather makes the same choices as long as her objective performance is above a certain threshold, viewed as an aspiration level. Our model follows similar lines in focusing on status quo decisions that aren't consciously, explicitly taken. As opposed to Simon's, our model does not attempt to model the choices made once a conscious decision is taken. On the other hand, the cases in our model need not be past choices by the same decision maker in a similar problem. In particular, our model more readily captures phenomena such as social learning, contagion, etc.

Kahneman (2010) suggested the distinction between the fast, intuitive, "System 1 (S1)", and the slow, more rational, "System 2 (S2)". Our distinction between unconscious and conscious choices would sometimes overlap the S1/S2 distinction. Following habits and sticking with status quo decisions could be viewed as belonging to S1, especially because the rational decision process of S2 isn't invoked in these cases. However, many of the examples of decisions made by S1 aren't status quo decisions. Correspondingly, S1 might lead to decisions that would be considered mistaken for various reasons, such as biases and heuristics. Of these, our focus is only on the status quo bias, that is, on situation in which the only possible mistake is refraining from thinking.

Cerigioni (2017) offers a model of sequential decision making, in which, as long as a decision problem is similar to past ones, the decision maker makes the same choice she made in the past, and she engages in conscious optimization only for novel problems. This is a behavioral model of the two systems, S1, S2, and their interaction. Like Cerigioni (2017), we are interested in a decision maker who sometimes makes a conscious decision and sometimes isn't. Moreover, choices that retain the status quo, such as keeping one's job or staying in one's country, can be viewed as making the same choice that has been made in similar cases in the past. However, there are several differences between the models. First, ours is silent on the decision making process once it is set into motion. Second, our model allows for various cases to invoke conscious decision making, and not only the appearance of novel problems.[1] Third, Cerigioni (2017) assumes that only choices are observable, and that the existence of a conscious decision process should be inferred from them. By contrast, our model discusses circumstances under which deliberation would occur, as if the existence of such deliberation were observable. Our model is not designed to identify the existence of S1 thinking from data, but to clarify what is actually assumed when status quo decisions are modeled as if they were conscious and rational.

2 Model and Representation

The set of case-types is C, assumed finite, with $|C| = n$. A database is a function $I : C \to \mathbb{Z}_+$, with $\mathcal{I} = \mathbb{Z}_+^n$ being the set of all databases. For $I \in \mathcal{I}$ define $|I| = \sum_{j=1}^{n} I(j)$. $0 \in \mathcal{I}$ denotes the origin, and summation, subtraction, multiplication by a number, and inequality on $\mathcal{I}$ are interpreted pointwise.

[1]This is similar to the distinction between our model and satisficing: while Simon (1957) suggests that a decision maker would be prompted into making a decision by low payoffs, Cerigioni (2017) assumes that it is the similarity, rather than the payoff function, that is key to deliberation. Both, however, focus on past choices by the same decision maker.

A set $\mathcal{S} \subset \mathcal{I}$ is given, interpreted as the set of databases for which the status quo is retained. The set $\mathcal{D} = \mathcal{I} \backslash \mathcal{S}$ is interpreted as the set of databases for which a decision will be consciously made. We assume

A1. Non-triviality: $0 \in \mathcal{S}$; $\mathcal{D} \neq \varnothing$.

Thus, A1 states that both $\mathcal{S}$ and $\mathcal{D}$ are not empty: there are databases for which the status quo will be chosen without deliberation, and there are others for which a conscious decision will be made. Moreover, in the absence of any data, the status quo will be retained.

Next, we assume the Combination condition of Gilboa and Schmeidler (2003) for the set $\mathcal{S}$, but not for $\mathcal{D}$. This condition states that, if two disjoint databases lead the decision maker keeping the status quo without deliberation, so should their union. In our formulation, the union is modeled by the sum of the two counter vectors that represent the databases:

A2. S-Combination: If $I, J \in \mathcal{S}$, then $I + J \in \mathcal{S}$.

The logic of this condition is the following: if a database I supports that conclusion that the current decision is the optimal one, and so does a database J, when they are taken together the conclusion can only be further reinforced. We do not assume a similar condition for the set $\mathcal{D}$, because it is possible that one database, I, gives reason to believe that an act a is better than the status quo, while another, J, that an act $b \neq a$ is better than the status quo, but given the union, $I + J$, neither a nor b seems very promising, and the status quo decision can still appear to be a reasonable choice, or a safe compromise.

By contrast, if we add a database I to itself, it stands to reason that, if we start in $\mathcal{D}$ we will also end up in it. Thus we have

A3. D-Replicability: If $I \in \mathcal{D}$, then $kI \in \mathcal{D}$ for $k \in \mathbb{N}$.

Notice that $\mathcal{D}$-Replicability could follow from a condition similar to the Combination one, if I were restricted to be a replication of J. Clearly,

Observation 1 *Given A2 and A3, for every $I \in \mathcal{I}$ and every $k \in \mathbb{N}$, $I \in \mathcal{S}$ iff $kI \in \mathcal{S}$ (and $I \in \mathcal{D}$ iff $kI \in \mathcal{D}$).*

We also adapt an Archimedean condition from Gilboa and Schmeidler (2003):

A4. Archimedeanity: If $I \in \mathcal{D}$, then for every $J \in \mathcal{S}$ there exists positive integer k, $kI + J \in \mathcal{D}$.

Archimedeanity states that, if database I contains sufficient information to set a decision making process in motion, then, even if database J does not, sufficiently many replications of I would do the same, that is, would also make the decision maker "wake up".

A *theory* $t : C \rightarrow (0, 1)$ assigns weight to case type. It is implicitly assumed that this weight is independent of past cases.

Theorem 1 *The following two statements are equivalent:*

(i) A1-A4 hold;

(ii) There exists a set of theories T with $t_0 \in T$, and there is a $t' \in T$ with $\frac{t_0(c)}{t'(c)} < 1$ for some $c \in C$, such that, for every $I \in \mathcal{I}$, $I \in \mathcal{S}$ iff

$$\Pi_c t_0(c)^{I(c)} \geq \Pi_c t(c)^{I(c)}, \ \forall t \in T.$$

Thus, conditions A1-A4 are equivalent to the representation we started out with, that is, to the claim that the decision maker behaves as if she were explicitly considering a set of theories about alternative paths of action, but dismisses them based on the data.

Note that the theorem says nothing about uniqueness. It will be obvious from the proof that the representation is not unique. We only comment here that, if we multiply $t(c)$ by a constant $a_c > 0$ for all theories t, the representation remains valid.

A natural question to ask is, can we interpret theories as probability distributions over C? The short answer is "almost". Indeed, our representation result $\Pi_c t_0(c)^{I(c)} \geq \Pi_c t(c)^{I(c)}$ is equivalent to $\Sigma_c I(c) ln \frac{t_0(c)}{t(c)} > 0$, leading us to think of $ln\frac{t_0(c)}{t(c)}$ as log-likelihood ratio. Let us define $r_t(c) = ln\frac{t_0(c)}{t(c)}$ for all t and c. Fix a $t \in T$, it is clear from the representation that the $r_t(c)$'s are unique up to multiplication by a common factor $k_t > 0$. To make the probabilistic interpretation rigorous, we need to find a probability $t_0 \in \Delta(C)$ such that $\Sigma_c \frac{t_0(c)}{exp(k_t r_t(c))} = 1$ for all $t \in T$, where $t_0(c)/exp(k_t r_t(c))$ is the probability theory t put on case c. If we set $t_0(c) = 1/n$ for all $c \in C$, it is not guaranteed that we can make every theory t a probability distribution over C. However, if we add an auxiliary case c_0 and set $t_0(c) = 1/(n+1)$ for all $c \in C \cup \{c_0\}$, we can pick $k_t > 0$ close enough to 0 such that $\Sigma_{c\in C} \frac{1}{exp(k_t \cdot r_t(c))} \leq n+1$. Then, we set $t(c) = t_0(c)/exp(k_t \cdot r_t(c))$ and $t(c_0) = 1 - \Sigma_{c\in C} t(c) \geq 0$. Now, all theories in T are probability distributions over $C \cup \{c_0\}$. What's left is to understand each database $I \in \mathcal{I}$ as if it contains 0 observation of case c_0.

3 Extensions

In the previous section we used two assumptions that, combined, implied that $I \in \mathcal{S}$ iff $kI \in \mathcal{S}$ for all I and all natural k. This property simplifies the analysis and, in particular, allows a relatively simple condition such as Combination to imply the convexity of the set $\mathcal{S}$. However, the assumption that all that matters is the relative frequency of case types in a database, and that its size has no import is somewhat restrictive. For example, a single case of a hate crime may be shrugged off as an aberration, but ten repetitions of such cases may set a decision process in motion. More generally, if we take into account statistical considerations, and wish to describe a decision maker who only "wakes up" in face of accumulated evidence, we need to relax these assumptions.

We first ask the question that when we can categorize all databases into two sets $\mathcal{S}'$ and $\mathcal{D}'$ that satisfy all the conditions in Theorem 1.

A set $\mathcal{Y} \subset \mathcal{I}$ is *segment-integer-convex* if, for every $I, J \in \mathcal{Y}$, and every $\alpha \in [0,1]$, if $\alpha I + (1-\alpha) J \in \mathcal{I}$, then $\alpha I + (1-\alpha) J \in \mathcal{Y}$.

A5. Convexity: $\mathcal{S}$ is segment-integer-convex.

A5 states that if two databases don't have enough evidence against the status-quo, their "weighted average" should not be able to overturn the status quo either.

Define $\mathcal{S}' = \{I \in \mathcal{I} : kI \in \mathcal{S}, for\ all\ k \in \mathbb{N}\}$ as the set of databases that do not have strong evidence against the status quo even after arbitrarily many replications. Let $\mathcal{D}' = \mathcal{I} \backslash \mathcal{S}'$. In the appendix we show that if A1, A4, and A5 hold for $\mathcal{S}$ and $\mathcal{D}$, then for any $J \in \mathcal{D}'$, there is a $I \in S$ such that $I + J \in D$, which serves as a characterization of $\mathcal{D}'$.

This would allow us to state the following.

Theorem 2 *If conditions A1, A4, and A5 hold for $\mathcal{S}$ and $\mathcal{D}$, then conditions A1, A2, A3, and A4 hold for $\mathcal{S}'$ and $\mathcal{D}'$.*

Theorem 2 says that A1, A4, and A5 allow us to unveil a set $\mathcal{S}'$ of databases that are "for the status quo" and a set $\mathcal{D}'$ of databases that are "against the status quo". What's more, $\mathcal{S}'$ and $\mathcal{D}'$ satisfy all the conditions required by Theorem 1, which leads to the following corollary.

Corollary 1 *If conditions A1, A4, and A5 hold for $\mathcal{S}$ and $\mathcal{D}$, then there exists a set of theories T with $t_0 \in T$, and there is a $t' \in T$ with $\frac{t_0(c)}{t'(c)} < 1$ for some $c \in C$, such that, for every $I \in \mathcal{I}$, $I \in \mathcal{S}'$ iff*

$$\Pi_c t_0(c)^{I(c)} \geq \Pi_c t(c)^{I(c)}, \ \forall t \in T.$$

A set $\mathcal{Y} \subset \mathcal{I}$ is *integer-convex* if, for every $m \geq 2$, every $(I_j)_{j=1}^m \subset \mathcal{Y}$, and $(\alpha_j)_{j=1}^m$ such that (i) $\alpha_j \geq 0$; (ii) $\sum_{j=1}^m \alpha_j = 1$; (iii) $\sum_{j=1}^m \alpha_j I_j \in \mathcal{I}$, then $\sum_{j=1}^m \alpha_j I_j \in \mathcal{Y}$. Notice that in this definition $(\alpha_j)_{j=1}^m$ may be assumed rational without loss of generality.

A5'. Convexity: $\mathcal{S}$ is integer-convex.

This would allow us to state the following.

Theorem 3 *If A1 and A5' hold, then there exists a set of theories T with $|T| > 1$ and $t_0 \in T$, and constants $(d_t)_{t\in T}$ in $(0,1]$ such that, for every $I \in \mathcal{I}$,*

$$I \in \mathcal{S} \ if \ \Pi_c t_0(c)^{I(c)} > d_t \Pi_c t(c)^{I(c)} \ for \ all \ t \in T$$
$$I \in \mathcal{D} \ if \ \Pi_c t_0(c)^{I(c)} < d_t \Pi_c t(c)^{I(c)} \ for \ some \ t \in T.$$

4 Discussion

4.1 Axioms and Conditions

Our results deal with characterizations of certain representations. We consider a set of databases for which the status quo is retained, $\mathcal{S}$, and provide conditions that are sufficient, or necessary and sufficient for the set to be described in a certain mathematical way. This type of exercise is often referred to as an "axiomatization". While axiomatizations are considered to be at the foundations of decision theory, one might ask, what is the point in axiomatizing, or characterizing the set $\mathcal{S}$?

One reason for the interest in axiomatic systems is the need to relate theoretical concepts to observations. For example, the abstract notion of "utility" can be given meaning by its relationship to observed choices. But in our case it is not obvious that the process we deal with, namely the decision maker "waking up" to make a conscious decision, is so easily observable.[2] Another common reason to be interested in axiomatizations is normative: axioms may convince decision makers that they would like to follow a certain decision procedure (such as utility maximization, expected utility maximization,

[2] In this sense, Cerigioni (2017) can be considered a more classical axiomatization, as his model assumes that only bottom-line choices are observable, and the two systems interpretation is the theoretical constructs that is presumable identified from the data.

and so forth). This, however, is hardly the case here: dealing with the implicit decision whether to make an explicit decision, there seems to be little point in trying to convince one's subconsciousness how it should make its decisions.

There is, however, another reason for which axioms are useful, and this is to help economists understand when various types of models, or conceptual paradigms make sense. For example, vNM's (1944) axiomatization of expected utility maximization under risk convinced economists that this model should be the model of choice also for descriptive, and not only for normative purposes. That is, the axioms were playing a rhetorical role, but their audience was not the decision maker, but the economist who modeled decision makers.

Our goal is along these lines, but it is rather modest. We do not argue that the conditions we state on the set $\mathcal{S}$ are necessarily very compelling, or that economists should use our representation results in future work. Rather, our goal is to clarify what is being assumed when one does not model implicit status quo decisions. An economist who would read these conditions and find them plausible would be justified in assuming that decision makers behave *as if* they made a conscious, explicit decision at every stage. By contrast, to the extent that these condition seem counter-intuitive, one may wish to delve deeper into the decision making process and ask what happens if and when individuals follow a status quo without making a conscious decision. For this reason we refer to our conditions as such, rather than as "axioms": we do not attempt to convince the reader that they are more or less universal principles; rather, we view them as tools that can aid in the economist's modeling decision, whether to adopt the as-if-conscious model.

4.2 Compatibility with the Bayesian Account

In our model the status quo decision is rationalized by a set of theories, of which one obtains the maximal likelihood value. There could be, however, many other ways of rationalizing status quo decisions. In particular, one may adopt a Bayesian account, according to which the decision maker starts out with a prior probability over all future observations, as well as over the outcomes all her acts might result in, and, given this prior and the observed database, she updates her beliefs and makes a decision that maximizes her (subjective, posterior) expected utility. In this model Mary's decision to keep her job can be viewed as if, every morning, she considered her career choices, and makes a decision that most of the time turns out to be to keep the status quo. Should there be any evidence that makes her change this choice, such as a negative review, it can be viewed as a signal processed by a rational decision maker in a Bayesian way, changing her posterior beliefs and thereby also her optimal choice.

This approach is both insightful and powerful. Indeed, if Mary stays on her job for years, one might conclude that it probably wasn't all that bad, considering what she believed the alternatives to be. Further, there is almost no real-life event that cannot be explained, at least post-hoc, using this rational, Bayesian approach. If, given a class databases, Mary does not make a conscious decision, one can imagine a prior belief that, conditional on each of these databases, leads to a posterior according to which the status quo promises higher subjective expected utility than do its alternatives.[3] Yet, we suspect that such an explanation might be too general and becloud some behavioral patterns. For example, suppose that career changes are contagious, in the sense that one is more likely to consider one's career when one's friends do so. Career changes made by others can be viewed as signals that induce a Bayesian updating of a person's beliefs about the alternatives available to her. This would indeed be a rather convincing account of contagion if the

[3]This probably holds for experimental data as well, when considering the information that might be implicit in the experiment's protocol. One caveat may be a version of the Sure Thing Principle, according to which, if, given any possible information that might be received a certain choice is optimal, so it should be before information is received. See Green and Park (1996), Shmaya and Yariv (2016).

person makes a similar career move to those made by her friends. Yet, if she makes a completely different career change, the information account might be somewhat artificial. There are situations where it would seem more natural to think of the decision maker "waking up" to consider her own problem, which might differ in many ways from those of her friends.

5 Appendix: Proofs and Related Analysis

Proof of Theorem 1:

We first present a lemma.

For any $x \in \mathbb{R}^n$, an algebraically open neighborhood around x is a set $E \subset \mathbb{R}^n$ such that for any $y \in \mathbb{R}^n$, there is an $\alpha \in (0,1)$, $ry + (1-r)x \in E$ for all $r \in [0, \alpha)$. Obviously, E is an algebraically open neighborhood around x if and only if we can associate each point e in the unit sphere a positive number ϵ_e such that $\{y \in \mathbb{R}^n : ||y-x|| < \epsilon_{(y-x)/||y-x||}\} \subset E$. Notice that if the ϵ_e's can be chosen uniformly over all directions, E is an open neighborhood around x.

Lemma : Let $E \subset \mathbb{R}^n$ be an algebraically open neighborhood around x. Let $F \subset \mathbb{R}^n$ be a convex set. If $E \cap F = \emptyset$, then $x \notin \bar{F}$.

Proof. Suppose there is a sequence $\{x_n\}$ in F converging to x. For any set $H \subset \mathbb{R}^n$, let $dim(H)$ be the dimension of the subspace spanned by the elements of H. Let $m = lim_{k \to +\infty} dim(\{x_n - x\}_{n \geq k})$. Clearly m exists and $1 \leq m \leq n$. We know $m > 1$, because $m = 1$ will imply that the x_n's eventually converge to x from a single direction, i.e., that for some k and for all $n > k$, $x_n - x$ are proportional, which will be in direct violation of the algebraic openness of E. Let k^* be large enough so that we have $m = dim(\{x_n - x\}_{n \geq k^*})$. Fix the subspace spanned by $\{x_n - x\}_{n \geq k^*}$ and pick $m+1$ elements $\{y_i\}_{i=1}^{m+1}$ in $\{x_n - x\}_{n \geq k^*}$ such that $conv(\{y_i\}_{i=1}^{m+1})$ has nonempty interior relative to this subspace with the usual metric topology. This can be done, otherwise $\{x_n - x\}_{n \geq k^*}$ have to be on a hyperplane passing through the origin (because $x_n - x$ converges to the origin) in this subspace, which means $dim(\{x_n - x\}_{n \geq k^*})$ has to be smaller than m. Consider the interior of $conv(\{y_i\}_{i=1}^{m+1})$ and pick an element y_0 and an open ball $B(y_0, \epsilon)$ of radius $\epsilon > 0$ around it in this interior. Due to the convexity of F, we have $x + B(y_0, \epsilon) \subset F$. By algebraic openness of E, there is an $\alpha \in (0,1)$ such that $x + \alpha y_0 \in E$. For every $n > k^*$, the point $z_n = y_0 + \frac{1-\alpha}{\alpha}(x - x_n)$ satisfies $(1-\alpha)x_n + \alpha(x + z_n) = x + \alpha y_0$. Since $x_n \to x$, it must be the case that

$z_n \in B(y_0, \epsilon)$ for large n. Then, for large n, both $x + z_n$ and x_n are in F. This along with the convexity of F contradict $x + \alpha y_0 \in E$. ∎

A set E is algebraically open if every point in E has an algebraically open neighborhood in E.

Corollary 2: Let $E \subset \mathbb{R}^n$ be an algebraically open set. If E^c is convex, then E is open.

Proof. Every point in E has an algebraically open neighborhood in E that is disjoint from $\overline{E^c}$. Hence, $E \cap \overline{E^c} = \emptyset$. ∎

Corollary 3: Given a convex set $F \subset \mathbb{R}^n$ and a point $x \notin F$. If for any $y \in F$, there is an $\alpha \in (0, 1)$ such that $ry + (1-r)x \notin E$ for all $r \in [0, \alpha)$. Then, x is not in the closure of F.

Proof. The argument used in the lemma only requires the existence of α for directions coming from set F. ∎

Two observations about the Archimedeanity condition are due here. First, the Archimedeanity property also holds for $J \in \mathcal{D}$. Because, for any $J \in \mathcal{D}$, if $J + I \in \mathcal{D}$, we are done. If $J + I \in \mathcal{S}$, we apply the Archimedeanity condition to $J + I$ and get $(J + I) + kI \in \mathcal{D}$ for some $k \in \mathbb{N}$.

Second, for any $I \in \mathcal{D}$ and any $J \in \mathcal{I}$, there exists positive integer K such that for all $k > K$, $kI + J \in \mathcal{D}$. If $k_1 < k_2 < k_3$ and both $k_1 I + J$ and $k_3 I + J$ are in $\mathcal{S}$, we must have $k_2 I + J = \alpha(k_1 I + J) + (1-\alpha)(k_3 I + J)$ for some $\alpha \in \mathbb{Q} \cap (0, 1)$. Then, there exists positive integers n and m, such that $\alpha = \frac{m}{n}$ thus $n(k_2 I + J) = m(k_1 I + J) + (n-m)(k_3 I + J)$. Since $\mathcal{S} - combination$ says $m(k_1 I + J) + (n-m)(k_3 I + J) \in \mathcal{S}$, by Observation 1 it must be the case that $k_2 I + J \in \mathcal{S}$. Successive applications of Archimedeanity show that there are infinitely many ks such that $kI + J \in \mathcal{D}$. So, by the above argument, we can't have infinitely many ks such that $kI + J \in \mathcal{S}$.

We now proceed to the proof of Theorem 1.

Proof: (ii) implies (i) is obvious. We thus focus on showing that (i) implies (ii).

We extend the databases from $\mathcal{I} = \mathbb{Z}_+^n$ to $\mathbb{Q}_+^n$. For any $I \in \mathbb{Q}_+^n$, find a $q \in \mathbb{Z}_+$ such that $qI \in \mathcal{I}$ and let $I \in \mathcal{D}$ if and only if $qI \in \mathcal{D}$. This is well defined because both $\mathcal{D}$ and $\mathcal{S}$ have the replicability property. The union of these extended two disjoint regions is $\mathbb{Q}_+^n$. We still call them $\mathcal{D}$ and $\mathcal{S}$ respectively.

By replicability, we have $I \in \mathcal{D}$ if and only if $qI \in \mathcal{D}$ for all $q \in \mathbb{Q}_+$.

Due to $\mathcal{S}$-combination, it must be true that for any $I, J \in \mathcal{S}$, $pI + qJ \in \mathcal{S}$ for all $p, q \in \mathbb{Q}_+$. Moreover, this implies for any $\{I_i\}_{i=1}^l \subset \mathcal{S}$, we have $\Sigma_{i=1}^l r_i I_i \in \mathcal{S}$ whenever $\Sigma_{i=1}^l r_i = 1$ and $r_i \in (0,1) \cap \mathbb{Q}$ for all $i = 1, 2, ..., l$.

The Archimedeanity condition implies that given any $I \in \mathcal{D}$ and $J \in \mathbb{Q}_+^n$, there is a $r^* \in (0,1) \cap \mathbb{Q}$ such that $rI + (1-r)J \in \mathcal{D}$ for all $r \in (r^*, 1) \cap \mathbb{Q}$.

We want to show that the closure of $\mathcal{S}$ in $\mathbb{R}_+^n$, denoted by $\bar{\mathcal{S}}$, is a closed convex cone. It is closed by definition and it is convex because it is also the closure of a convex set $conv(\mathcal{S})$, i.e., the convex hull of $\mathcal{S}$ in $\mathbb{R}_+^n$. To see it is a cone, assume that $I \in \bar{\mathcal{S}}$ is the limit of a sequence $\{I_n\}$ in $\mathcal{S}$ and let q be in $\mathbb{Q}_+$. Since $\{qI_n\}$ must converge to qI, we have $I \in \bar{\mathcal{S}}$ if and only if $qI \in \bar{\mathcal{S}}$ for all $q \in \mathbb{Q}_+$. By the closedness of $\bar{\mathcal{S}}$, it must be the case that $qI \in \bar{\mathcal{S}}$ for all $q \in \mathbb{R}_+$. Hence, $\bar{\mathcal{S}}$ is a closed convex cone.

Next we show $\bar{\mathcal{S}} \cap \mathcal{D} = \emptyset$. We have $\mathcal{D} \cap conv(\mathcal{S}) = \emptyset$, because every rational point in $conv(\mathcal{S})$ can be obtained as a convex combination of points in $\mathcal{S}$ using only rational coefficients, and thus every such rational point is in $\mathcal{S}$ and not in $\mathcal{D}$. Then, the result follows from Corollary 3 because $\mathbb{Q}$ is a field.

Given any $\beta \in \mathbb{R}^n$, let $c_\beta = inf_{x \in \bar{\mathcal{S}}} x \cdot \beta$. Because $\bar{\mathcal{S}}$ is a cone, it must be the case that $c_\beta \in \{-\infty, 0\}$. Let $B = \{\beta \in \mathbb{R}^n : ||\beta||_1 = 1, c_\beta = 0\}$, where $||\beta||_1 = \Sigma_{i=1}^n |\beta(i)|$. The supporting halfspace associated with a vector $\beta \in B$ is denoted by $H_\beta = \{x \in \mathbb{R}^n : x \cdot \beta \geq 0\}$. It's well known that $\bar{\mathcal{S}} = \cap_{\beta \in B} H_\beta$. Thus, we conclude that, for all $I \in \mathcal{I}$, I is in $\mathcal{S}$ if and only if, for every $\beta \in B$ we have $\Sigma_{c \in C} I(c)\beta(c) \geq 0$. Define a theory t_0 by $t_0(c) = 1/n$ for all $c \in C$.

For every $\beta \in B$, define a theory t_β by letting $t_\beta(c) = \frac{t_0(c)}{exp(\beta(c))} = \frac{1}{nexp(\beta(c))}$ for all $c \in C$. Because β is identified up to a positive scale factor, we can always choose the scale close to 0 in order to guarantee $t_\beta(c) \in (0,1)$ for all $c \in C$. Denote the collection of theories by $T = \{t_\beta\}_{\beta \in B'} \cup \{t_0\}$. Because $\Sigma_{c\in C} I(c)\beta(c) \geq 0$ if and only if $\Pi_{c\in C}[exp(\beta(c)]^{I(c)} \geq 1$, it follows that for all $I \in \mathcal{I}$, I is in $\mathcal{S}$ if and only if $\Pi_c t_0(c)^{I(c)} \geq \Pi_c t(c)^{I(c)}$ for all $t \in T$. ∎

Note that the set of theories need not be minimal. It's obvious that hyperplanes with all coefficients strictly positive (or strictly negative) won't have any bite. For hyperplanes with all coefficients nonpositive (or nonnegative), we only need to consider those with only one coefficient not being 0. What's more, redundancies may also appear as illustrated in the following example. Consider $\mathcal{S} = \{I \in \mathbb{Z}^2_+ : I(2) \geq I(1)\}$. The two essential supporting halfspaces are $\{I \in \mathbb{Z}^2_+ : I(1) \geq 0\}$ and $\{I \in \mathbb{Z}^2_+ : -I(1) + I(2) \geq 0\}$. However, $\{I \in \mathbb{Z}^2_+ : -I(1) + \alpha I(2) = 0\}$ with $\alpha > 1$ are all supporting halfspaces.

We present a stronger notion of Archimedeanity for reference.

A4'. Archimedeanity: If $J + I \in \mathcal{D}$ for some $J \in \mathcal{S}$, then for any $J' \in S$, there exists positive integer k such that $kI + J' \in \mathcal{D}$.

Remark: Define $\mathcal{S}' = \{I \in \mathcal{I} : kI \in \mathcal{S}, \textit{for all } k \in \mathbb{N}\}$. The A4' Archimedean condition is equivalent to the statement that given any database J in $\mathcal{S}$, $J + kI$ is in $\mathcal{S}$ for all $k \in \mathbb{N}$ if and only if I is in $\mathcal{S}$'.

We now proceed to the proof of Theorem 2.

We first present two observations. First, for any $I \in \mathcal{D}$, if $k_1 < k_2 < k_3$ and both $k_1 I + J$ and $k_3 I + J$ are in $\mathcal{S}$, segment-integer convexity of $\mathcal{S}$ implies that $k_2 I + J \in \mathcal{S}$. This along with the Archimedeanity condition, implies there can not exist infinitely many ks such that $kI + J \in \mathcal{S}$. Thus, there is a positive integer K such that for all $k > K$, $kI + J \in \mathcal{D}$.

Second, given $0 \in \mathcal{S}$, $\mathcal{D}$-replicability is implied by the segment-integer convexity of $\mathcal{S}$. If $I \in \mathcal{D}$ and $kI \in \mathcal{S}$ for some $k > 1$, then $I = \frac{k-1}{k}0 + \frac{1}{k}kI \in \mathcal{S}$, which is a contradiction.

Proof of Theorem 2: Recall that $\mathcal{S}' = \{I \in \mathcal{I} : kI \in \mathcal{S}, for\ all\ k \in \mathbb{N}\}$, and $\mathcal{D}' = \mathcal{I} \backslash \mathcal{S}'$. Obviously, $\mathcal{S}'$ and $\mathcal{D}'$ are disjoint and $\mathcal{S}' \cup \mathcal{D}' = \mathcal{I} = \mathbb{Z}_+^n$. What's more, we have $\mathcal{S}' \subset \mathcal{S}$ and $D \subset D'$.

We first give a characterization of $\mathcal{D}'$: for any $J \in \mathcal{D}'$, then there is a $I \in S$ such that $I + J \in D$. If $J \in \mathcal{D}$, then we use $I = 0 \in \mathcal{S}$. If $J \in \mathcal{S}$, the number $k(J) = max\{k \in \mathbb{Z}_+ : kJ \in \mathcal{S}\}$ is well defined because $0 \in \mathcal{S}$ together with the (segment-integer) convexity of $\mathcal{S}$ imply at most finitely many kJs are in $\mathcal{S}$. Then, for $I = k(J) \cdot J \in \mathcal{S}$ we have $I + J \in \mathcal{D}$.[4]

If $I \notin \mathcal{S}'$, then it must be the case that $kI \notin \mathcal{S}'$ for all $k > 1$ due to $0 \in \mathcal{S}$ and the (segment-integer) convexity of $\mathcal{S}$. Hence $\mathcal{D}'$ satisfies the counterpart of $\mathcal{D}$-Replicability: $I \in \mathcal{D}'$ implies that $kI \in \mathcal{D}'$ for all k.

If $I, J \in \mathcal{S}'$, then we have $kI, kJ \in \mathcal{S}$ for all k by definition, which, along with the (segment-integer) convexity of $\mathcal{S}$, leads to $k(I+J) = \frac{1}{2}2kI + \frac{1}{2}2kJ \in \mathcal{S}$ for all $k \in \mathbb{N}$. Hence, $I + J \in \mathcal{S}'$ and $\mathcal{S}'$ satisfies the counterpart of $\mathcal{S}$-Combination condition.

If $J \in \mathcal{D}'$, we have $kJ \in \mathcal{D}$ for some $k \in \mathbb{N}$. Then, for any $I \in \mathcal{I}$, the Archimedeanity condition says $I + nkJ \in \mathcal{D}$ for some $n \in \mathbb{N}$. So, we have $I + nkJ \in \mathcal{D}'$ because $D \subset D'$. This implies that $\mathcal{D}'$ and $\mathcal{S}'$ satisfy the counterpart of Archimedeanity condition. ∎

Proof of Corollary 1: Given Theorem 2, we can apply Theorem 1 to $\mathcal{S}'$ and $\mathcal{D}'$, leading to the claim. ∎

Proof of Theorem 3: For any $\beta \in \mathbb{R}^n$, let $c_\beta = inf_{I \in \mathcal{S}} \Sigma \beta(i) I(i)$. Clearly, every c_β is nonpositive because 0 is in $\mathcal{S}$. Let $B = \{\beta \in \mathbb{R}^n : ||\beta||_1 = 1, c_\beta \neq$

[4]If we have A4' instead of A4, the converse of this observation is also true and thus we have $\mathcal{D}' = \{I \in \mathcal{I} : \exists J \in \mathcal{S}, J + I \in \mathcal{D}\}$. This is because if we have $J + I \in \mathcal{D}$ for some $J \in \mathcal{S}$, then, by A4' applied for $J' = 0$, we can not have $kI \in S$ for all $k \in \mathbb{N}$, i.e. $I \notin \mathcal{S}'$ and thus $I \in \mathcal{D}'$.

$-\infty\}$. Then, we have $\overline{conv(\mathcal{S})} \cap \mathbb{Z}_+^n = \cap_{\beta \in B}\{x \in \mathbb{Z}_+^n : x \cdot \beta \geq c_\beta\}$.

Define a theory t_0 by $t_0(c) = 1/n$ for all $c \in C$. For every $\beta \in B$, define a parameter $d_\beta = exp(c_\beta)$ and define a theory t_β by letting $t_\beta(c) = \frac{t_0(c)}{exp(\beta(c))} = \frac{1}{nexp(\beta(c))}$ for all $c \in C$. Because (β, c_β) is identified up to a positive scale factor, we can always choose the scale close to 0 in order to guarantee $t_\beta(c) \in (0,1)$ for all $c \in C$. Denote the collection of theories by $T = \{t_\beta\}_{\beta \in B}$. Because $\Sigma_{c \in C} I(c)\beta(c) \geq c_\beta$ if and only if $\Pi_{c \in C}[exp(\beta(c)]^{I(c)} \geq exp(c_\beta)$ for all $\beta \in B$, it follows that $I \in \overline{conv(\mathcal{S})} \cap \mathbb{Z}_+^n$ if and only if $\Pi_c t_0(c)^{I(c)} \geq d_\beta \Pi_c t(c)^{I(c)}$ for all $t \in T$.

Suppose there is a $J \in \mathcal{D}$ and $J \in \overline{conv(\mathcal{S})} \cap \mathbb{Z}_+^n$. Look at the subspace in $\mathbb{R}^n$ spanned by $\mathcal{S}$, equipped with the relative topology. Since this subspace is closed, $\overline{conv(\mathcal{S})}$ is in this subspace. And, because $0 \in \mathcal{S}$, $conv(\mathcal{S})$ must have positive "volume" in this subspace, which implies that the interior of $\overline{conv(\mathcal{S})}$ in the relative topology is nonempty. Then, J must be disjoint from the interior of $\overline{conv(\mathcal{S})}$ in the relative topology, otherwise we can find point which belongs to $conv(\mathcal{S})$ in each orthant of a small open ball around J (think J as the origin), which contradicts the convexity of $\mathcal{S}$. By a hyperplane separation argument, there is a hyperplane in this subspace weakly separates J and $\overline{conv(\mathcal{S})}$. By extending this hyperplane to a hyperplane in the whole space $\mathbb{R}^n$, we conclude that for some $\beta \in B$ we have $c_\beta = \Sigma_i J(i)\beta(i)$.

Hence, the claim of the theorem is valid. ∎

Three examples in $\mathbb{Z}_+^2$ illustrate some technical problems:

Example 1: Take $\mathcal{S} = \{I \in \mathbb{Z}_+^2 : I(2) > \sqrt{2}(I(1) - 2)\}$. So the new Archimedean condition is satisfied and we have $\mathcal{S}' = \{I \in \mathbb{Z}_+^2 : I(2) > \sqrt{2}I(1)\}$. The essential supporting halfspaces of $\mathcal{S}'$ are $\{I(1) \geq 0\}$ and $\{-\sqrt{2}I(1) + I(2) \geq 0\}$. Because $\{n\sqrt{2} \ mod \ 1 : n \in \mathbb{N}\}$ is dense in $[0,1]$, it must be true that for any $\epsilon > 0$, there is a $I \in \mathcal{S}$ such that $I(1) - \sqrt{2}(I(1) - 2)\} \in (0, \epsilon)$. This implies $inf_{I \in \mathcal{S}}\{I(2) - \sqrt{2}I(1)\} = 2\sqrt{2}$ and there is no $I \in \mathcal{S}$ that achieve this lower bound. Then, notice that $(2,0) \in \mathcal{D}$ is on the hyperplane $I(2) = \sqrt{2}(I(1) - 2)$, i.e. some database in $\mathcal{D}$ sneaks in the supporting hyperplanes. Notice that this won't be a problem if we just characterize $\overline{conv(\mathcal{S})}$, since $(2, 0)$ is not in that closure. But, if the essential decision process is driven by the two theories revealed by $\mathcal{S}'$, characterizing $\overline{conv(\mathcal{S})}$ will introduce many extra theories whose existence is simply due to the discreteness of integers.

Example 2: Boundary 1 connects (0,0) to the ray starting from (2,1) with a 45 degree angle. Boundary 2 connects (0,0) to the ray starting from (1,2) with a 45 degree angle. Let $\mathcal{S}$ be the set of all the integer points on and between these two boundaries. $\mathcal{S}' = \{I \in \mathbb{Z}^2_+ : I(1) - I(2) = 0\}$ and the two essential supporting halfspaces of $\mathcal{S}'$ are $\{I \in \mathbb{Z}^2_+ : -I(1)+I(2) \leq 0\}$ and $\{I \in \mathbb{Z}^2_+ : I(1) - I(2) \leq 0\}$. But the essential supporting halfspaces of $\overline{conv(\mathcal{S})}$ also contain $\{I \in \mathbb{Z}^2_+ : I(1) - 2I(2) \leq 0\}$ and $\{I \in \mathbb{Z}^2_+ : 2I(1) - I(2) \leq 0\}$.

Example 3: $\mathcal{S}$ being unbounded does not imply $\mathcal{S}'$ contains a database other than 0. Let α be an irrational in $(\frac{1}{2}, 2)$. Boundary 1 connects (0,0) to the ray starting from (2,1) with slope α. Boundary 2 connects (0,0) to the ray starting from (1,2) with slope α. Let $\mathcal{S}$ be the set of all the integer points on and between these two boundaries. Because $\{n\alpha \ mod \ 1 : n \in \mathbb{N}\}$ is dense in $[0, 1]$, $\mathcal{S}$ must contain infinitely many databases. But, in this example we have $\mathcal{S}' = \{0\}$.

6 References

Akaike, H. (1974), "A new look at the statistical model identification". *IEEE Transactions on Automatic Control* **19** (6), 716–723.

Cerigioni, F. (2017), "Dual Decision Processes: Retrieving Preferences when some Choices are Automatic".

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

——— (2010), “Likelihood and Simplicity: An Axiomatic Approach”, *Journal of Economics Theory*, **145**: 1757-1775.

Green, E. J. and Park, I-U. (1996), “Bayes Contingent Plans”, *Journal of Economic Behavior and Organizations*, **31**: 225-236.

Kahneman, D. (2011), *Thinking, Fast and Slow.* Farrar, Straus, and Giroux.

March, J. G. and H. A. Simon (1958), *Organizations.* New York: Wiley.

Savage, L. J. (1954), *The Foundations of Statistics.* New York: John Wiley and Sons. (Second addition in 1972, Dover)

Shmaya, E., and L. Yariv (2016), “Experiments on Decisions Under Uncertainty: A Theoretical Framework”, *American Economic Review*, **106**: 1775-1801.

Simon, H. A. (1957), *Models of Man.* New York: John Wiley and Sons.

Sims, C. A. (2003), “Implications of Rational Inattention”, *Journal of Monetary Economics*, **50**: 665-690.

von Neumann, J. and O. Morgenstern (1944), *Theory of Games and Economic Behavior.* Princeton: Princeton University Press .