

A Graph-based Similarity Function for CBDT: Acquiring and Using New Information

Federico Contiggiani

Universidad Nacional de Río Negro
Instituto de Investigación en Políticas Públicas y Gobierno

Fernando Delbianco

Departamento de Economía, Universidad Nacional del Sur
Instituto de Matemática de Bahía Blanca
CONICET - Universidad Nacional del Sur

Fernando Tohmé

Departamento de Economía, Universidad Nacional del Sur
Instituto de Matemática de Bahía Blanca
CONICET - Universidad Nacional del Sur

August 30, 2020

Abstract

We present a formal model of decision-making under uncertainty, a variant of Case-Based Decision Theory, in which the solution to a problem obtains in terms of the distance to previous problems. We formalize this by defining a space based on an orthogonal basis of *features* of problems. New problems are evaluated in this setting, determining the best available actions in terms of the distance to previous problems. This is particularly relevant in the case of categorization and decision making, in which information systems have to support efficient ways of solving problems. We also show how this framework evolves upon the acquisition of new information, namely features or values of them arising in new problems. We discuss how this can be useful to evaluate decisions based on not yet existing data.

JEL Classification: D01, D81.

Keywords: Microeconomic Behavior, Decision-Making under Risk and Uncertainty, Case Based Decision Theory.

1 Introduction

Cognitive models of human decision making have enriched economic theory by providing relevant insights on the psychological and computational process that operate when the individual evaluates alternatives and makes a decision (Lipman, 1995). Behavioral approaches to individual decision making lend support to the metaphor that human choices might be well represented as an information processor that is influenced by elements of the decision context (Rubinstein, 1998).

Information processing is critically dependent on the accessibility to the right sources and the accuracy of data. This is even more relevant when this information is required for decision-making under uncertainty. Some important questions in this respect are, *What information is needed to solve a problem?*, *How can the relevant factors be weighted in the search of solutions?* and *How can new aspects in that search be learned or inferred?* Recent contributions show that these questions are still open and call for new concepts and methodologies (Newell and Bröder, 2008; Lipshitz et al., 2001).

Our take on this subject is based on a particular model of Case-Based Decision Theory (CBDT). This approach assumes that the similarity with previous problems is key in the way a new problem is addressed. Similar approaches in judgment and decision making (JDM) literature refer to *exemplar* models for the study of categorization strategies (Karlsson, Juslin, and Olsson, 2008), cue-based inferences elaboration and example-based reasoning (Platzer and Bröder, 2013). In addition there are methodological questions common to both CBDT and Naturalistic Decision Making (Lipshitz et al., 2001).

The canonical approach of individual decision making under uncertainty is Expected Utility Theory (EUT), which assumes that rational agents enumerate all the possible states of the world and the corresponding consequences associated with them. Furthermore, they are able to assess the probabilities of all possible relevant states of the world. In the real world, the agents are usually unable to fully describe the class of states of the world, mainly because of the complexity involved in this task.

Instead, when individuals face decision problems, they make their decisions based on their previous experience, searching in their memories to recall what they did in similar situations, in order to assess the convenience of choosing the same actions as in the past. This intuition is captured by CBDT, as presented by Gilboa and Schmeidler, 1995; Gilboa and Schmeidler, 2001; Gilboa and Schmeidler, 2003 and Matsui, 2000. In CBDT the preferences of a decision-maker over actions to be exerted to solve a new problem result from the history of previous problems faced by her, which are stored in her memory.

A few assumptions ensure that those preferences can be captured in an utility function over actions, defined as the sum of values of adopting them on previous problems. Each of these values is weighted up by the similarity between the previous and current problem. Similarities are so crucial in this framework that we can identify an agent at a new problem with the ensuing similarity function. The original formulation of CBDT does not assume any particular

shape of this function. But in other areas, analogous relations have been characterized (Tversky, 1977; Lipman, 1995; Rubinstein, 1998; Mullainathan, 2002; Kahneman, 2003; Johnson and Ratcliff, 2014).

According to this, we can define similarity in terms of a “distance” between problems, which can be obtained, in turn, by comparing the features characterizing them. The closer two problems are in terms of such distance, the higher is their similarity. In our framework we construct a graph embedded in a real metric space. Each vertex is identified with a problem and the edges connecting pairs of problems have an attached weight, namely the distance between them. When a new problem arises, it becomes a point in the metric space. The problems in the graph that are *closer* to it become relevant to determine the actions to be carried out. Learning amounts to reconfigure the metric space and the embedded graph upon the acquisition of new pieces of information.

The metric space in which the graph of problems is embedded is defined by the *features* that constitute the dimensions on which the similarities of problems are evaluated. This is close to the idea that decisions are made in terms of those made previously by peers (Li, Sun, and Ettredge, 2017). Furthermore, we evaluate how decisions are made in the presence of new values of the features or in the face of the realization of the relevance of new features. Furthermore, we consider how the anticipation of future new values or features affect the decisions.

In this paper we present the formal aspects of this decision-making framework, starting with a thorough description of CBDT. We add an extra axiom that ensures the characterization of similarity in terms of the distance between problems. Then we introduce the metric space capturing this concept of distance and analyze how it evolves in response to new information.

The work is structured as follows. Section 2 presents the formalism of CBDT. Section 3 discusses how to represent CBDT’s similarity function in terms of distances on graphs. Section 4 specifies those graphs in spaces determined by *features* of problems. Section 5 studies how the previously presented formalism can be adapted to new information. In turn, Section 6 discusses how agents can learn the rate at which new data arises and use this to make decisions about contexts that are not yet existing. Section 7 concludes.

2 Definitions and Axioms

Gilboa and Schmeidler (1995) assume a finite and nonempty set P which is given as a primitive and contains all the possible *problems* that an agent may face as well as a finite and nonempty set of *actions* A . To simplify they assume that all the actions in A are available for any problem $p \in P$. In addition, there is a set $R = \mathbb{R}$ of results. The result of not choosing an action is denoted $r_0 \in R$ (for simplicity, we assume that $r_0 = 0$). Then, the set of *cases* is $C \equiv P \times A \times R$. The agent is endowed with a *memory* set $M \subseteq C$. Its projection over P is called the *history* and is defined by $H(M) = \{q \in P \mid \exists a \in A, r \in R : (q, a, r) \in M\}$. The set M has the following two properties:

- for all $q \in H(M) \wedge a \in A \implies \exists! r = r_M(q, a) : (q, a, r) \in M$;
- for all $q \in H(M) \implies \exists! a \in A : r_M(q, a) \neq r_0$.

The first condition indicates that for every problem q in the history and an action a there exists a unique result of applying this action to solve q . The second condition states that for each problem q in the history, there exists a unique action a that yields a non-void result. These two conditions together ensure that for each pair of cases in H , (q, a, r) and (q', a', r') , $q \neq q'$ and $a \neq a'$.

The agent makes a decision based both on the utility of the results of a given action, and a *similarity* function s which assigns nonnegative values to pairs of problems. In this way, when the agent faces a new problem p , he selects an act a that maximizes the following expression:

$$U(a) = U_{p,M}(a) = \sum_{(q,a,r) \in M} s(p, q) u(r)$$

where $u : R \rightarrow \mathbb{R}^+$ is the *instantaneous utility* of results.

Gilboa and Schmeidler present an axiomatic system, representing some desirable properties of the similarity function. Furthermore, they show that there exists a unique $s : P^2 \rightarrow [0, 1]$ such that the function $U(\cdot)$ is the representation of a preference relation $\succeq_{p,H}$, on the class of actions, where p is a new problem not corresponding to any case in the history. In the statement of the axioms each action $x \in A$ is identified with a vector in \mathbb{R}^H , where each component of the vector is the result of applying action x on a problem q for which a case (q, a, r) exists in the history. We denote with $x(q)$ the result of applying action x to problem q . If $x = a$, then $x(q) = r$. Otherwise, $x(q) = r_0$.

Axiom 1 *Comparability of Compatible Profiles.* For every $p \in P$ and every history $H = H(M)$, for every $x, y \in \mathbb{R}^H$, x and y are compatible if and only if $x \succeq_{p,H} y$ or $y \succeq_{p,H} x$.

Axiom 2 *Monotonicity.* For every p, H , $x \geq y$ and $x * y = 0$ implies that $x \succeq_{p,H} y$.

Axiom 3 *Continuity.* For every p, H , and $x \in \mathbb{R}^H$, the sets $\{y \in \mathbb{R}^H | y \succeq_{p,H} x\}$ and $\{y \in \mathbb{R}^H | x \succeq_{p,H} y\}$ are closed (in the standard topology on \mathbb{R}^H).

Axiom 4 *Separability.* For every p, H , and $x, y, z, w \in \mathbb{R}^H$, if $(x + z) * (y + w) = 0$, $x \succeq_{p,H} y$, and $z \succeq_{p,H} w$, then $(x + z) \succeq_{p,H} (y + w)$.

Axiom 5 *Similarity Invariance.* For every $p, q_1, q_2 \in P$ and every two memories M^1, M^2 with $q_1, q_2 \in H^i \equiv H(M^i)$ ($i=1,2$) and $p \notin H^i$ ($i=1,2$), let v_j^i stand for the unit vector in \mathbb{R}^{H^i} ($i=1,2$) corresponding to q_j ($j=1,2$) (That is, v_j^i is a vector whose q_j th component is 1 and its other components are 0). Then denoting the symmetric part of $\succeq_{p,H}$ by $\approx_{p,H}$,

$$x, y \in \mathbb{R}^{H^1}, z, w \in \mathbb{R}^{H^2}, x \approx_{p,H^1} y, z \approx_{p,H^2} w$$

and

$$x + \alpha v_1^1 \approx_{p,H^1} y + \beta v_2^1$$

imply that

$$z + \alpha v_1^2 \approx_{p,H^2} w + \beta v_2^2$$

whenever the compared profiles are compatible.

Given these axioms and using claims proven by Gilboa and Schmeidler, we can show that the similarity between pairs of problems is representable in terms of a connected graph G in which each problem in P is assigned a node. The class of nodes of G is V and thus $V \equiv P$. To get there, we need first to consider two extra axioms.

Axiom 6 *Symmetric Similarity.* For every $r, p, m \in P$ and every three memories M^1, M^2, M^3 with $m, r \in H^1 \equiv H(M^1)$, $p \notin H^1$, $m, p \in H^2 \equiv H(M^2)$, $r \notin H^2$, $p, r \in H^3 \equiv H(M^3)$, and $m \notin H^3$; let v_j^i stand for the unit vector in \mathbb{R}^{H^i} ($i=1,2,3$) corresponding to j ($j=r,p,m$) (That is, is a vector whose j th component, which is the component associated with the case j , is 1 and its other components are 0). Then denoting the symmetric part of $\succeq_{p,H}$ by $\approx_{p,H}$,

$$x, y \in \mathbb{R}^{H^1}, z, w \in \mathbb{R}^{H^2}, l, h \in \mathbb{R}^{H^3}, x \approx_{p,H^1} y, z \approx_{r,H^2} w, l \approx_{m,H^3} h$$

and

$$x + \alpha v_m^1 \approx_{p,H^1} y + v_r^1, z + \beta v_m^2 \approx_{r,H^2} w + v_p^2$$

imply that

$$l + \alpha v_p^3 \approx_{m,H^3} h + \beta v_r^3$$

whenever the compared profiles are compatible.

Axiom 6 guarantees that the similarity function is symmetric since it leads us to Gilboa and Schmeidler's necessary and sufficient condition for symmetric similarity, namely that for all $r, p, m \in P$, $s(p, m) s(m, r) s(r, p) = s(p, r) s(r, m) s(m, p)$. More precisely,

Theorem 1 *The following two statements are equivalent:*

- Axiom 1 to Axiom 6 hold.*
- There exist a unique and symmetric function $s' : P^2 \rightarrow \mathbb{R}_+$ such that for all $p \in P$, every memory M with $p \notin H \equiv H(M)$ and every compatible $x, y \in \mathbb{R}^H$, $x \succeq_{p,H} y \iff \sum_{q \in H} s'(p, q) x(q) \geq \sum_{q \in H} s'(p, q) y(q)$.*

Now we introduce another axiom that requires that the similarity function verifies the following inequality, i.e. $s'(p, q) \geq s'(p, m) + s'(m, q)$. Intuitively, this amounts to ask that the "direct" similarity between two problems cannot be lower than the sum of the similarities of these problems mediated by another one. More precisely:¹

¹Notice that this axiom, as shown through its consequences in Theorem 2, precludes the possibility of, given three problems p, m and q , to define a similarity function such that $s(p, q) = s(p, m) = s(m, q)$.

Axiom 7 *Triangular inequality.* For every $r, p, m \in P$ and every two memories M^1, M^2 with $m, r \in H^1 \equiv H(M^1)$, $p \notin H^1$, $m, p \in H^2 \equiv H(M^2)$, and $r \notin H^2$; let v_j^i stand for the unit vector in \mathbb{R}^{H^i} ($i=1,2$) corresponding to j ($j=r,p,m$) (That is, v_j^i is a vector whose j th component, which is the component associated with the case j , is 1 and its other components are 0). Then denoting the symmetric part of $\succeq_{p,H}$ by $\approx_{p,H}$,

$$x, y \in \mathbb{R}^{H^1}, z, w \in \mathbb{R}^{H^2}, x \approx_{p,H^1} y, z \approx_{r,H^2} w$$

and

$$x + \alpha v_m^1 \approx_{p,H^1} y + v_r^1, z + \beta v_p^2 \approx_{r,H^2} w + v_m^2$$

imply that

$$y + \beta v_r^1 + \alpha v_m^1 \leq_{p,H^1} x + v_m^1$$

whenever the compared profiles are compatible.

The addition of this axiom leads to:

Theorem 2 *The following two statements are equivalent:*

- a) *Axiom 1 to Axiom 7 hold.*
- b) *There exist a unique and symmetric function $s' : P^2 \rightarrow \mathbb{R}_+$ that verifies that $s'(p, q) \geq s'(p, m) + s'(m, q)$, such that for all $p \in P$, every memory M with $p \notin H \equiv H(M)$ and every compatible $x, y \in \mathbb{R}^H$,*

$$x \succeq_{p,H} y \iff \sum_{q \in H} s'(p, q) x(q) \geq \sum_{q \in H} s'(p, q) y(q).$$

The proof of both theorems is given in the Appendix.

3 The Similarity Function Represented by a Graph

Given Axioms 1 to 7, and the results obtained in the previous section, we are now in position to introduce a graph-theoretic version of the similarity function. For this, consider an agent endowed with a memory set M and a connected graph G with nodes $V = P$.

We assume that the agent is able to compute a distance $d(p, q)$ between $p, q \in V$, which is defined as the length of the shortest path joining these two nodes, and since G is connected, this distance is a metric (Harary, 1969).

In this context, *length* is defined as the number of occurrences of edges in an alternating sequence of nodes and edges (*walk*) between p and q (Harary, 1969). In this way, the agent is able to compare any pair of problems in the graph. Consider the adjacency matrix $B = [b_{pq}]$ of G that is the $n \times n$ matrix ($|P| = n$) in which $b_{pq} = 1$ if p is adjacent with q in G and $b_{pq} = 0$ otherwise. Then, the distance between p and q for $p \neq q$ is the least integer l for which the (p, q) , entry of B^l is nonzero.

Now, we are going to consider the following results.

Proposition 1 *Consider $S_G = \langle V, d \rangle$, where V is the set of nodes of graph G and d is Harary's minimal distance defined on G . Then, there exists $\rho : \ell \rightarrow S_G \cup \mathbb{R}^+$, where ℓ are all the expressions in the formal language in which*

the CBDT axioms are formulated, such that $M_G = \langle S_G \cup \mathbb{R}^+, \rho \rangle$ is a model of CBDT+Axiom 6+Axiom 7.

Proposition 2 *CBDT+Axiom 6+Axiom 7 has only one model up to isomorphism.*

Given these propositions we show that S_G provides a valid representation of the memory and the similarity function because expressions that are true of s are true in G with distance d ; and in addition, G is compatible with axioms 1 to 7. Furthermore, the representation of s using a graph G is unique up to isomorphism. Finally, all the expressions on utilities are interpreted as arithmetical claims in \mathbb{R}^+ . The proofs of these propositions can be found in the Appendix.

In other words, these propositions show that there exists a unique graph that represents the similarity function. This is an important addition to the usual presentation of CBDT, since the similarity function is derived there from the preferences over outcomes, while here it is furthermore associated to the shortest path distance on a graph.

We will now consider a concrete definition of graph G and derive an explicit specification of s .

4 A Specification of the Similarity Function

In order to present a concrete specification of the results of the previous section we need to define a graph G in which the nodes correspond to problems. One way to do that is by assuming that each problem can be identified to a point in a *space of features* \mathcal{F} . More precisely, since we intend to find a way of maximizing a preferential order $\succeq_{p,H}$, we are only interested in a space that represents the features of the problems corresponding to the finite number of cases in H . By the conditions on H , all these problems are different, and thus each will yield a different point in the space of features.

Suppose that a problem $q \in H(M)$ is defined by j_q features.² Thus, the total number of features to consider is $J = \sum_{q \in H(M)} j_q$. While it is clear that some features are correlated to others, we will assume without much loss of generality that the features are independent. This is the case when J is the minimal number of features needed to describe any problem $q \in H(M)$. This simplification allows us to consider that the J dimensions of the space of features are orthogonal.

The remaining question is what spaces correspond to the different features. While some features admit discrete values others require continuous ones. Since we intend to use Harary's distance, we need to be able to define the *adjacency* between problems, and thus the range of values of each feature has to be discrete.

²If the problem amounts to, say, choosing a smartphone, we can consider that the relevant features are the price, the brand, the size of its memory, the quality of the camera, etc. While large, the number of features at the moment of making a decision, is finite.

We assume thus that we identify each feature f_j with its discrete range.³ That is, $f_j = \{f_1^j, \dots, f_{\kappa_{f_j}}^j\}$, where κ_{f_j} is the (finite) number of possible values of f_j . Notice that this means that f_j is a linearly ordered set. Then, the entire space of features is $\mathcal{F} = \prod_{j=1}^J f_j$.

G will consist of the points $H(M) \subseteq \mathcal{F}$. We take then as edges all the linear L^1 (“taxicab” or “Manhattan”) segments between the points in \mathcal{F} , with the proviso that the distance between two consecutive values in a feature, say f_k^j and f_{k+1}^j is 1. Harary’s procedure gives us the *minimal distance* between points on \mathcal{F} . Given two problems $q, r \in \mathcal{F}$, the ensuing distance according to this procedure is $d(q, r)$. In turn, the maximal distance between any pair of problems in $H(M)$ (the so called “diameter” of graph G) is denoted D_M .

Then, given two problems $p \in H(M)$ and $q \in H(M)$, we take $\bar{d}(p, q) = \frac{d(p, q)}{D_M}$ and define

$$s(p, q) = 1 - \bar{d}(p, q)$$

It is easy to see that:

Proposition 3 $s(p, q)$ satisfies the conditions of the similarity function of Theorem 2. Furthermore, for any statement about the similarity between any pair of problems p and q , $\Lambda(p, q)$, $\rho(\Lambda(p, q)) = s(p, q)$, where ρ is as defined in Proposition 1.

Example 3 Consider the following context:

$$H(M_{\mathcal{F}_0}) = \{q_i\}_{i=1}^4$$

where each problem is identified with an element of $\mathcal{F}_0 = f_1 \times f_2$, the space of features of the problems in memory, all related to buying or not cellphones. Here f_1 is the space of screen sizes, f_2 is the space of RAM memory sizes:

- $f_1 = \{5, 5.5, 7\}$,
- $f_2 = \{16, 32\}$,

The corresponding graph has diameter $D_{M_{\mathcal{F}_0}} = 3$.

A case $c_i \in M$ associated to a problem q_i , $i = 1, \dots, 4$, is then described as:

$$c_i = \langle q_i, a_i, r_i \rangle \in \mathcal{F}_0 \times \{0, 1\} \times R$$

where a_i is the decision made (either “buy” or “not buy”) while $r_i \in R$ is the result, understood as a degree of satisfaction (a real number in the interval $[0, 10]$).

³In the case of a continuous-valued feature this means that we determine a finite partition of its range. In practical applications the partition would arise from the application of methods like, for instance, CART.

The cases are:

$$\begin{aligned}c_1 &= \langle (5, 16), 1, 5 \rangle \\c_2 &= \langle (5.5, 16), 0, 10 \rangle \\c_3 &= \langle (5, 32), 0, 7 \rangle \\c_4 &= \langle (5.5, 32), 1, 7 \rangle\end{aligned}$$

Now suppose a new problem $p \in \mathcal{F}_0$ appears, namely to buy or not a configuration $(7, 16)$. The corresponding distances to the problems in $H(M_{\mathcal{F}_0})$ are: $d(p, q_1) = 2$, $d(p, q_2) = 1$, $d(p, q_3) = 3$ and $d(p, q_4) = 2$. Then (see Figure 1),

$$\begin{aligned}s(p, q_1) &= 1 - \frac{2}{3} = \frac{1}{3} \\s(p, q_2) &= 1 - \frac{1}{3} = \frac{2}{3} \\s(p, q_3) &= 1 - \frac{3}{3} = 0 \\s(p, q_4) &= 1 - \frac{2}{3} = \frac{1}{3}\end{aligned}$$

Then, the agent has to choose between $a = 1$ (“buy”) and $a = 0$ (“not buy”). The corresponding preferences are represented as (we assume $u(r) = r$ for every $r \in R$):

$$\begin{aligned}U(1) &= s(p, q_1)u(5) + s(p, q_4)u(7) = \\&= \frac{1}{3} \times 5 + \frac{1}{3} \times 7 = 4 \\U(0) &= s(p, q_2)u(10) + s(p, q_3)u(7) = \\&= \frac{2}{3} \times 10 + 0 \times 7 = 6\frac{2}{3}\end{aligned}$$

Since $U(0) > U(1)$ the decision is not buy the object.

5 Acquiring New Information

Up to this point, we have taken the entire M as source for the similarity relation. But new information may appear that could require to revise the decisions made

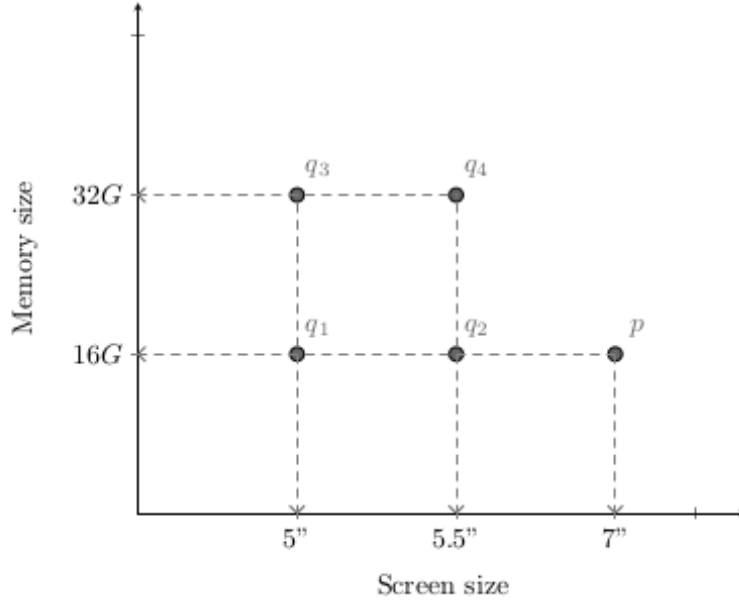


Figure 1: Graph representation of CBDT.

previously and change the way of addressing new problems. There are two instances that we want to consider:

- (a) A problem p that may appear with a new set of salient features $\{f_{1^p}, \dots, f_{m^p}\}$. Each f_j , $j = 1^p, \dots, m^p$ is either included in a space orthogonal to \mathcal{F} and to each f_k for $k \neq j$ or there exists a feature of \mathcal{F} , say f_l , such that $f_l \subset f_j$, i.e. f_j adds extra values to feature f_l .
- (b) p may not be compared to any other problem in $H(M)$ but only to some aspect of them.

Case (a) presents two subcases:

- (a1) A feature incorporates a new value.
- (a2) A new feature becomes relevant.

In these two instances the relevant space becomes $\mathcal{F}^p \times \prod_{k=1}^m f_{k^p}$, where \mathcal{F}^p are either the new features or the ones already in \mathcal{F} with the new values detected in p . Then, a new $s(\cdot, \cdot)$, based on the corresponding distance in the new graph, must be computed.

Example 4 (Case (a1)): consider the same M as in Example 3, where $H(M) = f_1 \times f_2 = \{5, 5.5\} \times \{16, 32\}$. Recall that $M = \{c_1; c_2; c_3; c_4\}$ is such that:

$$\begin{aligned}
q_1 &= (5, 16) \\
q_2 &= (5.5, 16) \\
q_3 &= (5, 32) \\
q_4 &= (5.5, 32)
\end{aligned}$$

with corresponding distances: $d(q_1, q_2) = 1$, $d(q_2, q_3) = 2$, $d(q_3, q_4) = 1$, $d(q_1, q_4) = 2$, $d(q_1, q_3) = 1$ and $d(q_2, q_4) = 1$. Then, since the diameter of the graph is $D_M = 2$, the similarity function is such that: $s(q_1, q_2) = \frac{1}{2}$, $s(q_2, q_3) = 0$, $s(q_3, q_4) = \frac{1}{2}$, $s(q_1, q_3) = \frac{1}{2}$, $s(q_2, q_4) = \frac{1}{2}$ and $s(q_1, q_4) = 0$ (see Figure 2).

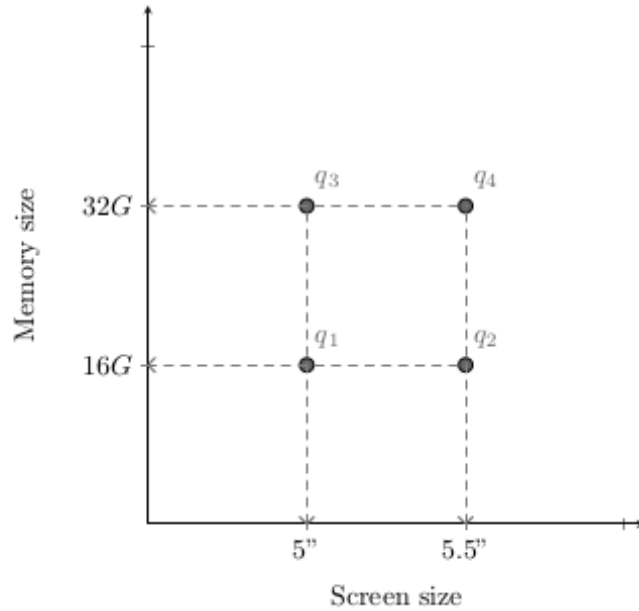


Figure 2: Example of $H(M)$ in Case (a1).

Now assume a new problem p in which feature f_1 presents a new value, 7. In the new graph, corresponding to $M' = M \cup \{p\}$, while the distances between q_1, \dots, q_4 are the same, the similarity function changes since the diameter of the graph is now $D_M = 3$. We have that: $s(q_1, q_2) = \frac{2}{3}$, $s(q_2, q_3) = \frac{1}{3}$, $s(q_3, q_4) = \frac{2}{3}$, $s(q_1, q_3) = \frac{2}{3}$, $s(q_2, q_4) = \frac{2}{3}$ and $s(q_1, q_4) = \frac{1}{3}$ (see, again, Figure 1).

Example 5 (Case (a2)): Assume again M as in Example 3, but now a new feature f_3 becomes relevant. Then, problems q_1, \dots, q_4 have to be redefined, acquiring a new component. That is, $q_i = \langle q_i^{f_1}, q_i^{f_2} \rangle$ becomes $q'_i = \langle q_i^{f_1}, q_i^{f_2}, \bar{q}_i^{f_3} \rangle$,

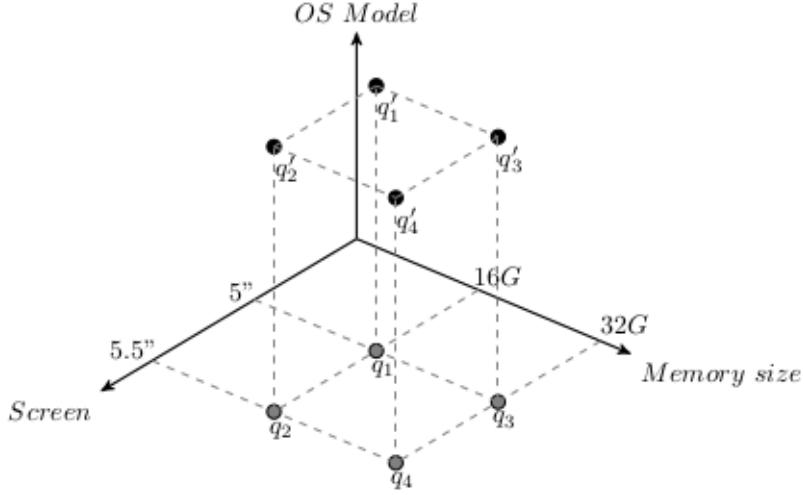


Figure 3: Example of Case (a2)

where $\bar{q}_i^{f_3} \in f_3$. in case $\bar{q}_i^{f_3}$ is not known with precision or it is not defined for q_i (for instance, the operating system of a an old cellphone), it is assigned an arbitrary value in f_3 . We represent this in Figure 3, in which the redefined problems get a non-null value in the third coordinate.

Case (b) poses a different question, namely to find aspects that are shared with previous problems. It can be addressed assuming that, given a new problem p , it comes associated to a subspace \mathcal{F}_0 of \mathcal{F} and a distance $\delta > 0$. For each problem $q \in H(M)$ we can define $q|_{\mathcal{F}_0}$, its projection over \mathcal{F}_0 . Then $H(M)^p = \{q \in H(M) : s(p|_{\mathcal{F}_0}, q|_{\mathcal{F}_0}) > \delta\}$, will be the class of problems to be taken into account for the choice of the optimal action. In case that $H(M)^p = \emptyset$, then by default we consider the entire $H(M)$.

Example 6 (Case (b)): Consider again M as in Example 3 and the problem of buying or not a phone $p = (7, 32, 9)$, but only in terms of the comparison with the features $\mathcal{F}_0 = f_1 \times f_2$. Then, for each $q_i \in H(M)$, $q_i|_{\mathcal{F}_0} = q_i$ while $p|_{\mathcal{F}_0} = (7, 32)$. Figure 4 represents this case, where the similarity function is:

$$\begin{aligned}
s(p, q_1) &= 1 - \frac{3}{3} = 0 \\
s(p, q_2) &= 1 - \frac{2}{3} = \frac{1}{3} \\
s(p, q_3) &= 1 - \frac{2}{3} = \frac{1}{3} \\
s(p, q_4) &= 1 - \frac{1}{3} = \frac{2}{3}
\end{aligned}$$

If $\delta = \frac{1}{2}$, $H(M)^p = \{q_4\}$ and thus, since $c_4 = (q_4, 1, 7)$, the decision should be “buy”.

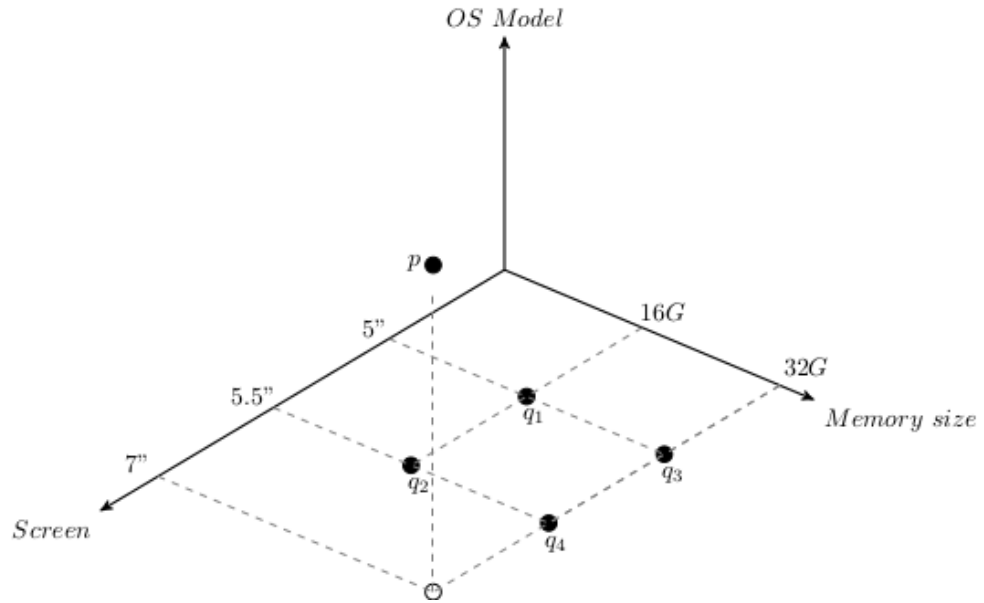


Figure 4: Example of Case (b)

6 Learning

The question becomes now if agents can anticipate new features or values of them, and consequently make choices taken those forecasts into account. This can only be probabilistically defined, and requires some assumptions. The first one we will make is that all the distributions are Poisson, as to capture the intuition that the relevant events (new features and values) happen at a rate independent of what happened at the previous period. This assumption can

be criticized on different grounds, but disregard it amounts to accept either a stronger regularity (a certain “law” of evolution of features and values) or a weaker one, making harder to forecast future events.

The second, and perhaps more critical, assumption is that the intervals defining the Poisson distributions represent the time interval between the formulation and solution of two different problems. While it does not need to be of a fixed duration, it tends to have a certain regularity in itself.

Let us define what it means for values and features, in terms of cases **(a1)** and **(a2)** (case **(b)** is of different nature and will not be treated here):

- (a1)** Any new problem p has a probability $Prob(f_{jp}) \sim Poisson(\lambda_j)$ of adding a new value to any feature f_j , for a given parameter λ_j . Furthermore, we assume that $\lambda_j = \lambda_k$ for any pair of features f_j and f_k already present in $H(M)$.
- (a2)** A new feature appears in any problem p with probability $Prob(J) \sim Poisson(\lambda_J)$.

The parameters λ represent the average number of changes in either the amount of values in a feature or in the number of features per relevant interval. To determine them, we have to compute some values. Given any new problem p we denote with Δ_{jp} the amount of new values of feature f_j and by Δ_{Jp} the amount of new features in p .

Then, given two memories M and M' , we define

$$\lambda_j = \frac{\sum_{p \in M' \setminus M} \Delta_{jp}}{|M' \setminus M|} \quad \text{and} \quad \lambda_J = \frac{\sum_{p \in M' \setminus M} \Delta_{Jp}}{|M' \setminus M|}$$

In either case the probability of k new values or features is given by

$$f(k, \lambda) = \frac{e^{-\lambda} \lambda^k}{k!}$$

The assessment of the values of λ_j and λ_J proceeds by progressive refinement, reevaluating those parameters starting from $M_0 = \emptyset$ and computing them for certain sizes of the memory. That is, every some fixed number of problems, the parameters are calculated again.

This process of continuous updating does not necessarily converges to fixed distributions. Agents will use the distributions they have at hand at the moment of making their decision. Furthermore, since it would be in anticipation of events that will happen at some time in the future, they have to *discount* the future outcomes to make meaningful comparison with decisions to be made in present time.

We can consider lotteries over a space $\mathcal{M} = \{(M, t, p) : M \text{ is a memory, } t \text{ is the moment at which it is assumed to be available and } p \text{ is a problem for which a solution is sought}\}$. Assuming preferences \succeq_{vNM} over \mathcal{M} , satisfying the usual *von Neumann-Morgenstern* axioms, we can represent them by an expected utility \bar{U} . It is immediate that $\bar{U}((M, t, p))$ corresponds to a lottery in which the

probability of $\langle M, t, p \rangle$ is 1 and the probability of every other $\langle M', t', p' \rangle \in \mathcal{M}$ is 0. The value of $\bar{U}(\langle M, t, p \rangle)$ can be identified with $\kappa^{t-t_0} U_{p,M}(a^*)$ where a^* is the action a that maximizes $\sum_{(q,a,r) \in M} s(p,q)u(r)$ and t_0 is the moment at which the lotteries are evaluated.⁴

Example 7 : Consider again M as in Example 3 and the problem of buying or not a phone $p = (7, 32, 9)$ now (period t) or wait until $t + 2$ to buy a better phone with a fourth feature f_4 and one higher value in features f_2 and f_3 . Since $t + 2$ is two intervals away, in terms of Poisson's distribution, the probabilities of those events are: $f(1, 2\lambda_j)$, for $j = 2, 3$ and $f(1, 2\lambda_J)$.

Taking $\lambda_j = \frac{1}{2}$ and $\lambda_J = \frac{1}{20}$, we have that the probabilities of new (higher) values in f_2 and f_3 are both $\frac{1}{e}$, while that of a new feature f_4 is $\frac{1}{10 e^{10}}$.

The similarities between p and the problems in $H(M)$ are ($D_M = 4$):

$$\begin{aligned} s(p, q_1) &= 1 - \frac{4}{4} = 0 \\ s(p, q_2) &= 1 - \frac{1}{4} = \frac{3}{4} \\ s(p, q_3) &= 1 - \frac{3}{4} = \frac{1}{4} \\ s(p, q_4) &= 1 - \frac{2}{4} = \frac{2}{4} \end{aligned}$$

While that of the potential p' are (the new M' has a diameter⁵ $D_{M'} = 7$):

$$\begin{aligned} s(p', q_1) &= 1 - \frac{7}{7} = 0 \\ s(p', q_2) &= 1 - \frac{6}{7} = \frac{1}{7} \\ s(p', q_3) &= 1 - \frac{6}{7} = \frac{1}{7} \\ s(p', q_4) &= 1 - \frac{5}{7} = \frac{2}{7} \end{aligned}$$

Consider the decision of buying a cellphone. We can compare two lotteries. One yields the utility of buying p in t , denoted $U_p^t(1)$, with certainty.

The other lottery yields the expected utility of buying p' in $t + 2$, $U_{p'}^{t+2}(1)$, with probability $f(1, 2\lambda_j) \times f(1, 2\lambda_j) \times f(1, 2\lambda_J) = \frac{1}{e} \times \frac{1}{e} \times \frac{1}{10 e^{10}} = \frac{1}{10 e^{12}}$.

The first lottery yields:

$$U_p^t(1) = s(p, q_1) \times 5 + s(p, q_4) \times 7 = \frac{7}{2} = 3\frac{1}{2}$$

⁴This is consistent with the two levels of inductive reasoning in learning and induction presented in (Gilboa and Schmeidler, 2001; Gilboa and Schmeidler, 2003).

⁵In this case, the diameter is the maximal distance between q_1 and p'

while for the second lottery, we need to compute its expected utility, by taking into account the aforementioned probabilities as well as the discount rate:⁶

$$\begin{aligned} U_{p'}^{t+2}(1) &= \frac{1}{10} \frac{1}{e^{12}} (s(p', q_1) \times \kappa \times 5 + s(p', q_4) \times \kappa \times 7) = \\ &= \frac{1}{5} \frac{1}{e^{12}} \kappa \end{aligned}$$

Then, the decision to postpone buying a cellphone, waiting for a potential p' brand, makes only sense if $\kappa \geq 17.5 e^{12}$, i.e. the future utility must be exceedingly large to compensate the extremely low chances of obtaining p' ($\sim 6 \times 10^{-7}$).

7 Conclusions

We have presented a graph-based definition of similarity to be used in the framework of Case-Based Decision Theory. This allows us to compute easily the actions that should be performed to maximize utilities based on comparing problems with others solved in the past.

Our characterization allows to represent the acquisition of new information and capture the learning process of agents in time. Our results indicate that while no convergence can be ensured the assessment at some point of time allows to compare current and future decisions.

This is particularly relevant in the context of judgment and decision-making in categorization tasks, providing an optimization answer to the intuitions and demands for methodological advances made by Newell and Bröder, 2008. Further work involves refining this approach and running experiments to compare with the decisions actually made by human agents.

A Proof of Theorem 1.

We are going to show only that a) implies b) since the converse is trivial. Gilboa and Schmeidler (1995) showed that Axioms 1 to 5 are equivalent to the existence of a function $s : P^2 \rightarrow [0, 1]$ such that for all $p \in P$, every memory M with $p \notin H = H(M)$, and every compatible $x, y \in \mathbb{R}^H$, $x \succeq_{p,H} y \iff \sum_{q \in H} s(p, q) x(q) \geq \sum_{q \in H} s(p, q) y(q)$.

Now we are going to show that considering axiom 6 the function $s(p, \cdot)$ can be transformed in a symmetric function, i.e. we are going to show that there exist a scalar δ_p such that we can rescale $s(p, \cdot)$, separately for each p , to convert

⁶Notice that it has to be applied as an *interest rate*, improving the future value of the rewards.

it in a symmetric function. Thus, it suffices to show that there exist a function $s'(p, \cdot)$ and a scalar δ_p such that:

$$s'(p, q) = \delta_p s(p, q)$$

and that this function is going to verify:

$$s'(p, q) = s'(q, p)$$

From axiom 6,

$$x, y \in \mathbb{R}^{H^1}, z, w \in \mathbb{R}^{H^2}, l, h \in \mathbb{R}^{H^3}$$

$$\bullet \quad x \approx_{p, H^1} y \iff$$

$$\begin{aligned} \sum_{q \in \{H^1 \setminus \{m, r\}\}} s_{H^1}(p, q) x(q) + s_{H^1}(p, m) x(m) + s_{H^1}(p, r) x(r) = \\ \sum_{q \in \{H^1 \setminus \{m, r\}\}} s_{H^1}(p, q) y(q) + s_{H^1}(p, m) y(m) + s_{H^1}(p, r) y(r) \end{aligned}$$

$$\bullet \quad z \approx_{r, H^2} w \iff$$

$$\begin{aligned} \sum_{q \in \{H^2 \setminus \{m, p\}\}} s_{H^2}(r, q) z(q) + s_{H^2}(r, m) z(m) + s_{H^2}(r, p) z(p) = \\ \sum_{q \in \{H^2 \setminus \{m, p\}\}} s_{H^2}(r, q) w(q) + s_{H^2}(r, m) w(m) + s_{H^2}(r, p) w(p) \end{aligned}$$

$$\bullet \quad l \approx_{m, H^3} h \iff$$

$$\begin{aligned} \sum_{q \in \{H^3 \setminus \{p, r\}\}} s_{H^3}(m, q) l(q) + s_{H^3}(m, p) l(p) + s_{H^3}(m, r) l(r) = \\ \sum_{q \in \{H^3 \setminus \{p, r\}\}} s_{H^3}(m, q) h(q) + s_{H^3}(m, p) h(p) + s_{H^3}(m, r) h(r) \end{aligned}$$

and

$$\bullet \quad x + \alpha v_m^1 \approx_{p, H^1} y + v_r^1 \iff$$

$$\sum_{q \in \{H^1 \setminus \{m, r\}\}} s_{H^1}(p, q) x(q) + s_{H^1}(p, m) [x(m) + \alpha] + s_{H^1}(p, r) x(r) =$$

$$\sum_{q \in \{H^1 \setminus \{m, r\}\}} s_{H^1}(p, q) y(q) + s_{H^1}(p, m) y(m) + s_{H^1}(p, r) [y(r) + 1]$$

- $z + \beta v_m^2 \approx_{r, H^2} w + v_p^2 \iff$

$$\sum_{q \in \{H^2 \setminus \{m, p\}\}} s_{H^2}(r, q) z(q) + s_{H^2}(r, m) [z(m) + \beta] + s_{H^2}(r, p) z(p) =$$

$$\sum_{q \in \{H^2 \setminus \{m, p\}\}} s_{H^2}(r, q) w(q) + s_{H^2}(r, m) w(m) + s_{H^2}(r, p) [w(p) + 1]$$

then

$$s_{H^1}(p, m) \alpha = s_{H^1}(p, r) \text{ and } s_{H^2}(r, m) \beta = s_{H^2}(r, p)$$

imply that

- $l + \alpha v_p^3 \approx_{m, H^3} h + \beta v_r^3 \iff$

$$\sum_{q \in \{H^3 \setminus \{p, r\}\}} s_{H^3}(m, q) l(q) + s_{H^3}(m, p) [l(p) + \alpha] + s_{H^3}(m, r) l(r) =$$

$$\sum_{q \in \{H^3 \setminus \{p, r\}\}} s_{H^3}(m, q) h(q) + s_{H^3}(m, p) h(p) + s_{H^3}(m, r) [h(r) + \beta]$$

then

$$s_{H^3}(m, p) \alpha = s_{H^3}(m, r) \beta$$

whenever the compared profiles are compatible.

Therefore, axiom 6 implies that

$$s_{H^3}(m, p) s_{H^1}(p, r) s_{H^2}(r, m) = s_{H^3}(m, r) s_{H^2}(r, p) s_{H^1}(p, m)$$

Using axiom 5 in the last expression we obtain:

$$s(m, p) s(p, r) s(r, m) = s(m, r) s(r, p) s(p, m)$$

$$s(m, p) \delta_m = \delta_p s(p, m)$$

$$s'(m, p) = s'(p, m)$$

■

B Proof of Theorem 2.

We are going to show that a) implies b), the converse is trivial. From axiom 7,

- $x \approx_{p,H^1} y \iff$

$$\sum_{q \in \{H^1 \setminus \{m,r\}\}} s_{H^1}(p,q) x(q) + s_{H^1}(p,m) x(m) + s_{H^1}(p,r) x(r) =$$

$$\sum_{q \in \{H^1 \setminus \{m,r\}\}} s_{H^1}(p,q) y(q) + s_{H^1}(p,m) y(m) + s_{H^1}(p,r) y(r)$$

- $z \approx_{r,H^2} w \iff$

$$\sum_{q \in \{H^2 \setminus \{m,p\}\}} s_{H^2}(r,q) z(q) + s_{H^2}(r,m) z(m) + s_{H^2}(r,p) z(p) =$$

$$\sum_{q \in \{H^2 \setminus \{m,p\}\}} s_{H^2}(r,q) w(q) + s_{H^2}(r,m) w(m) + s_{H^2}(r,p) w(p)$$

and

- $x + \alpha v_m^1 \approx_{p,H^1} y + v_r^1 \iff$

$$\sum_{q \in \{H^1 \setminus \{m,r\}\}} s_{H^1}(p,q) x(q) + s_{H^1}(p,m) [x(m) + \alpha] + s_{H^1}(p,r) x(r) =$$

$$\sum_{q \in \{H^1 \setminus \{m,r\}\}} s_{H^1}(p,q) y(q) + s_{H^1}(p,m) y(m) + s_{H^1}(p,r) [y(r) + 1]$$

- $z + \beta v_p^2 \approx_{r,H^2} w + v_m^2 \iff$

$$\sum_{q \in \{H^2 \setminus \{m,p\}\}} s_{H^2}(r,q) z(q) + s_{H^2}(r,p) [z(p) + \beta] + s_{H^2}(r,m) z(m) =$$

$$\sum_{q \in \{H^2 \setminus \{m,p\}\}} s_{H^2}(r,q) w(q) + s_{H^2}(r,p) w(p) + s_{H^2}(r,m) [w(m) + 1]$$

then

$$s_{H^1}(p,m) \alpha = s_{H^1}(p,r) \text{ and } s_{H^2}(r,p) \beta = s_{H^2}(r,m)$$

imply that

- $y + \beta v_r^1 + \alpha v_m^1 \preceq_{p, H^1} x + v_m^1 \iff$

$$\sum_{q \in \{H^1 \setminus \{m, r\}\}} s_{H^1}(p, q) y(q) + s_{H^1}(p, m) [y(m) + \alpha] + s_{H^1}(p, r) [y(r) + \beta] \leq$$

$$\sum_{q \in \{H^1 \setminus \{m, r\}\}} s_{H^1}(p, q) x(q) + s_{H^1}(p, m) [x(m) + 1] + s_{H^1}(p, r) x(r)$$

then

$$s_{H^1}(p, m) \alpha + s_{H^1}(p, r) \beta \leq s_{H^1}(p, m)$$

whenever the compared profiles are compatible.

Therefore, axiom 7 implies that

$$s_{H^1}(p, m) \frac{s_{H^1}(p, r)}{s_{H^1}(p, m)} + s_{H^1}(p, r) \frac{s_{H^2}(r, m)}{s_{H^2}(r, p)} \leq s_{H^1}(p, m)$$

Using axioms 5 and 6 we obtain:

$$s'(p, m) \geq s'(p, r) + s'(r, m)$$

■

C Proof of Propositions.

Proof of Proposition 1.

The interpretation satisfies the following statements:

$$\rho^1) p \rightarrow \bar{p}$$

$$\rho^2) s'(p, q) \rightarrow d(\bar{p}, \bar{q})$$

$$\rho^3) \geq_s \rightarrow \leq_d$$

$$\rho^4) u(a) \rightarrow \bar{u}_a.$$

where $p, q \in P$, while $\bar{p}, \bar{q} \in V$ and $\bar{u}_a \in \mathbb{R}$.

$d(\cdot)$ is a metric; that is for all $\bar{p}, \bar{q}, \bar{m} \in V(G)$,

1. $d(\bar{p}, \bar{q}) \geq 0$, with $d(\bar{p}, \bar{q}) = 0$ if and only if $p = q$
2. $d(\bar{p}, \bar{q}) = d(\bar{q}, \bar{p})$
3. $d(\bar{p}, \bar{m}) \leq d(\bar{p}, \bar{q}) + d(\bar{q}, \bar{m})$;

and that the similarity function satisfies

- 1'. $s'(p, q) \geq 0$
- 2'. $s'(p, q) = s'(q, p)$
- 3'. $s'(p, m) \geq s'(p, q) + s'(q, m)$

Therefore,

$$s'(p, q) \geq s'(p, m) \implies d(\bar{p}, \bar{q}) \leq d(\bar{p}, \bar{m}).$$

That is, d is an interpretation of s' , while $V(G)$ is an interpretation of the problems in memory plus the current problem. Then, S_G provides an interpretation of both the memory set and the similarity relation. To see that $S_G \cup \mathfrak{A}$ yields a true interpretation of CBDT + Axiom 6 + Axiom 7, notice that the main aim of axioms 1 to 4 is to show that the similarity function combined with the U-maximization is derivable from observed preferences. Therefore, since that is settled, now we have to show that the $d(\cdot)$ defined over G satisfies axioms 5 to 7.

Axiom 5 implies that, given two nonempty sets $H, H^0 \subseteq P \setminus \{p\}$, and $m, q \in H, H^0$, then

$$\frac{s_H(p, q)}{s_H(p, m)} = \frac{s_{H^0}(p, q)}{s_{H^0}(p, m)}.$$

Since $d(\cdot)$ does not depend on the history because G contains all the $p \in P$ then trivially satisfies this axiom.

Axiom 6 implies that the similarity function is symmetric, i.e. that $s'(m, p) = s'(p, m)$. Since G is a connected graph then $d(\cdot)$ is a metric. Therefore, it is symmetric.

Axiom 7 implies that the similarity function satisfies $s'(p, m) \geq s'(p, q) + s'(q, m)$ implying that $d(\cdot)$ is a metric, since it satisfies the triangular inequality.

■

Proof of Proposition 2.

Notice that given a particular problem p^* , there exists a partition of P , denoted $\{C_i\}_{i \in N}$ (where N is the set of natural numbers) in which for each $i \in N$, $C_i = \{q \in P : s'(p^*, q) = i\}$.

Since any theory that determines a partition of a countable set in countable classes is categorical (Keisler, 1977), there exists a single model for s' up to isomorphism. Furthermore, since there exists a single, up to linear transformations (i.e. isomorphisms), assignation of values of cardinal utilities into the real numbers, there exists a single model for CBDT + Axiom 6 + Axiom 7, namely M_G .

■

References

- Gilboa, I. and D. Schmeidler (1995). “Case-Based Decision Theory”. In: *The Quarterly Journal of Economics* 110.3, pp. 605–639. ISSN: 0033-5533. DOI: 10.2307/2946694.
- Gilboa, Itzhak and David Schmeidler (2001). *A Theory of Case-Based Decisions*. Cambridge: Cambridge University Press. ISBN: 9780511493539. DOI: 10.1017/CB09780511493539.
- (2003). “Inductive Inference: An Axiomatic Approach”. In: *Econometrica* 71.1, pp. 1–26. ISSN: 0012-9682. DOI: 10.1111/1468-0262.00388.
- Harary, Frank (1969). *Graph theory*. Addison-Wesley Publishing Company, p. 274. ISBN: 9780201027877.
- Johnson, Eric J. and Roger Ratcliff (2014). “Computational and Process Models of Decision Making in Psychology and Behavioral Economics”. In: *Neuroeconomics*, pp. 35–47. DOI: 10.1016/B978-0-12-416008-8.00003-6.
- Kahneman, Daniel (2003). “Maps of Bounded Rationality: Psychology for Behavioral Economics”. In: *American Economic Review* 93.5, pp. 1449–1475. ISSN: 0002-8282. DOI: 10.1257/000282803322655392.
- Karlsson, Linnea, Peter Juslin, and Henrik Olsson (2008). “Exemplar-based inference in multi-attribute decision making: Contingent, not automatic, strategy shifts?” In: *Judgment and Decision Making* 3.3, p. 244.
- Keisler, H.Jerome (1977). “Fundamentals of Model Theory”. In: *Studies in Logic and the Foundations of Mathematics* 90, pp. 47–103. ISSN: 0049-237X. DOI: 10.1016/S0049-237X(08)71098-X.
- Li, Xudong (Daniel), Lili Sun, and Michael Ettredge (2017). “Auditor selection following auditor turnover: Do peers’ choices matter?” In: *Accounting, Organizations and Society* 57, pp. 73–87. ISSN: 0361-3682. DOI: 10.1016/J.AOS.2017.03.001.
- Lipman, Barton L. (1995). “Information Processing and Bounded Rationality: A Survey”. In: *The Canadian Journal of Economics* 28.1, p. 42. ISSN: 00084085. DOI: 10.2307/136022.
- Lipshitz, Raanan et al. (2001). “Taking stock of naturalistic decision making”. In: *Journal of behavioral decision making* 14.5, pp. 331–352.
- Matsui, Akihiko (2000). “Expected utility and case-based reasoning”. In: *Mathematical Social Sciences* 39.1, pp. 1–12. ISSN: 0165-4896. DOI: 10.1016/S0165-4896(99)00008-6.
- Mullainathan, S. (2002). “A Memory-Based Model of Bounded Rationality”. In: *The Quarterly Journal of Economics* 117.3, pp. 735–774. ISSN: 0033-5533. DOI: 10.1162/003355302760193887.
- Newell, Ben and Arndt Bröder (2008). “Cognitive processes, models and metaphors in decision research”. In: *Judgment and Decision Making* 3.3, p. 195.
- Platzer, Christine and Arndt Bröder (2013). “When the rule is ruled out: Exemplars and rules in decisions from memory”. In: *Journal of Behavioral Decision Making* 26.5, pp. 429–441.
- Rubinstein, Ariel. (1998). *Modeling bounded rationality*. The MIT Press, p. 208. ISBN: 9780262181877.

Tversky, Amos (1977). "Features of similarity." In: *Psychological Review* 84.4, pp. 327–352. ISSN: 0033-295X. DOI: 10.1037/0033-295X.84.4.327. URL: <http://content.apa.org/journals/rev/84/4/327>.