Evaluating Case-based Decision Theory: Predicting Empirical Patterns of Human Classification Learning

Andreas Duus Pape
and Kenneth J. Kurtz*

January 17, 2012

Abstract

We introduce a computer program which calculates an agent's optimal behavior according to Case-based Decision Theory (Gilboa and Schmeidler 1995) and use it to test CBDT against a benchmark set of problems from the psychological literature on human classification learning (Shepard, Hovland, and Jenkins 1961). This allows us to evaluate the efficacy of CBDT as an account of human decision-making on this set of problems.

We find: (1) The choice behavior of this program (and therefore Case-based Decision Theory) correctly predicts the empirically observed relative difficulty of problems in the benchmark human data, which is a strong vote of confidence in its favor. (2) 'Similarity' (how CBDT decision makers extrapolate from memory) is decreasing in Euclidean vector distance, consistent with evidence in psychology (Shepard 1987). (3) Average similarity is rejected in favor of additive similarity. (4) CBDT learns the correct solutions unrealistically fast relative to human learners.

Keywords: Case-based Decision Theory, Human Cognition, Learning, Agent-based Computational Economics, Psychology, Cognitive Science

JEL codes: D83, C63, C88

*Pape: Binghamton University (SUNY) Department of Economics. Corresponding author. apape@binghamton.edu, (607)777-2660. Kurtz: Binghamton University (SUNY) Department of Psychology.

1 Introduction

We present a computational implementation of Case-based Decision Theory (Gilboa and Schmeidler 1995) called the Case-based Software Agent or CBSA. CBSA is a computer program that calculates an agent's optimal behavior according to Case-based Decision Theory for an arbitrary problem. Like Expected Utility Theory, Case-based Decision Theory is a mathematical model of choice under uncertainty. Case-based Decision Theory has the following primitives: A set of *problems* or circumstances that the agent faces; a set of *actions* that the agent can choose in response to these problems; and a set of *results* which occur when an action is applied to a problem. Together, a problem, action, and result triplet is called a *case,* and can be thought of as one complete learning experience. The agent has a finite set of cases, called a memory, which it consults when making new decisions.

The Case-based Software Agent is a *software agent*, i.e. "an encapsulated piece of software that includes data together with behavioral methods that act on these data (Tesfatsion 2006)." CBSA computes choice data consistent with an instance of CBDT for an arbitrary choice problem or game, provided that the problem is well-defined and sufficiently bounded (see Section 3). To examine CBSA's relationship with human learning, we chose to generate data for a benchmark set of problems from the psychological literature on human classification learning starting with Shepard, Hovland, and Jenkins (1961). In these problems, human decision makers sort objects described by vectors of characteristics, such as color and shape, and are rewarded when they sort objects correctly. The data include variables such as probability of error over time, which allows one to observe relative difficulty and speed of problem-solving/learning. We generate simulated choice data that is both consistent with Case-based Decision Theory and directly comparable to human data collected on the same benchmark set of problems.

This is the first instance of simulating nuanced choice data implied by an economic decision theory that can be brought to existing human choice data from psychology. Behavioral economics is typically characterized by applying insights from psychology to economics via a mathematical model, sometimes with modification, which is then tested using economic statistical methodologies on economic data (e.g. Laibson (1997), Fudenberg and Levine (2006)). This paper does the reverse: it tests a decision theory from economics using psychological methods and data.

In the framework of Case-based Decision Theory, the effects of actions on new problems are extrapolated from memory by evaluating the *similarity* between problems. The extrapolation from problems in CBDT is similar to generalization from stimuli in psychology. The study of generalization in psychology has lead to a remarkably specific empirical estimate of the functional form of similarity among humans (Shepard 1987). We test this proposition with CBSA and find results consistent with Shepard's.

Case-based Decision Theory's Axiom A5, "Similarity Invariance," requires that perceived similarity does not vary over time or experience. This axiom is satisfied by CBSA. In the study of human classification learning by Nosofsky and Palmeri (1996), human subjects were presented with classification problems in which different characteristics of the objects are difficult to distinguish, which is the context in which similarity invariance would most likely be satisfied. This is the benchmark human data against which we compare CBSA.

First, we find the choice behavior of CBSA (and therefore Case-based Decision Theory) to be psychologically valid with regard to the relative difficulty of problems in a manner consistent with the human choice data. This consistency with human behavior should be taken as a vote of confidence in support of CBSA as

an account of human decision-making.

Second, we find that, consistent with research in psychology cited above (Shepard 1987), similarity functions that are decreasing in vector distance induce the best match to human data; some alternative similarity functions are rejected in their favor.

Third, we find that additive similarity provides a better match for human data than does average similarity. In additive similarity, the main form of similarity proposed by Gilboa and Schmeidler, utility of an act encodes estimates of payoff and frequency that the act has been chosen, while average similarity, an alternative also proposed by Gilboa and Schmeidler, utility only encodes estimates of payoff.

Fourth, we find that CBSA does not correctly predict the absolute speed of solving these problems in that CBSA makes fewer mistakes than human learners.

Our results suggest that the fit of CBDT to human data can be improved upon via two means. First, the addition of an endogenous similarity function which selects which characteristics of problems to identify as important, what psychologists call 'attentional focusing.' (That is, "Similarity Invariance" might be too restrictive to describe human behavior; relaxations of this axiom such as described in Gilboa, Lieberman, and Schmeidler (2006) may be more realistic.) Second, that there should be constraints on the speed of learning, such as forgetfulness or probabilistic instead of deterministic choice.

Below, we review relevant literature in economic decision theory and agent-based computational economics (Section 2); we define CBSA precisely, then show that CBSA correctly and completely implements Case-based Decision Theory (CBDT), and therefore an instance of CBSA is an instance of CBDT (Section 3); we then introduce the psychology of human classification learning and relate it to decision theory (Section 4). Then we empirically demonstrate our primary results described above, and discuss the implications for decision theory (Section 5).

2 Related Literature in Economics

Decision Theory. To define a decision theory in the tradition of Savage and von Neumann and Morgenstern, an agent's choice behavior is observed and, if the choice behavior follows certain axioms, then a mathematical representation of utility, beliefs, et cetera can be constructed (von Neumann and Morgenstern 1944, Savage 1954). In an implementation, this is turned on its head: the mathematical representation is taken as given and choice behavior is produced. The purpose is to generate choice behavior for particular problems in hopes of finding empirical patterns in choice behavior that were not be *a priori* obvious from the mathematical representation alone. These patterns, coupled with human choice data, can empirically test hypotheses that the implemented decision theory describes human behavior. We do that here.

Case-based Decision Theory (Gilboa and Schmeidler 1995)—hereafter, CBDT—postulates that when an agent is confronted with a new problem, she asks herself: How similar is today's case to cases in memory? What acts were taken in those cases? What were results? She then forecasts payoffs of actions using her memory, and choses the action with the highest forecasted payoff.

Formally, in CBDT, the following objects are taken as primitives: a finite set of problems $\mathcal{P}$, a set of results $\mathcal{R}= \mathbb{R}$, which are utility payoffs, and a finite set of acts $\mathcal{A}$ which interact with problems to form results. There exists $r_0 \in \mathcal{R}$, where $r_0 = 0$ and is defined as the value of an action not taken. A set of cases $\mathcal{C}$ is defined $\mathcal{C} = \mathcal{P} \times \mathcal{A} \times \mathcal{R}$, and a subset of cases $\mathcal{M} \subseteq \mathcal{C}$ will be called the "memory" of this agent. The

memory is the data set that the agent draws on to make decisions. Under the choice axioms specified in Gilboa and Schmeidler (1995), a similarity function $s(p,q)$, $p,q \in \mathcal{P}$ can be specified such that the utility forecast, called case-based utility or CBU, is:

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

Where $M(a)$ is defined as the subset of the agents' memory $\mathcal{M}$ in which action a was taken.[1] This utility represents the agent's preference in that, for a fixed memory M, a is strictly preferred to a' if and only if $CBU(a) > CBU(a')$.

Comparing CBDT to Expected Utility Theory (EUT), one finds the most notable distinction is that EUT sums over states, which are exogenously given, while CBDT sums over memory, which is endogenous to agent experience. Similarity roughly serves the role that probability has in EUT, in that it weighs results to form a forecast. Similarity is weakly positive like probability, but is not constrained to sum to one. That probability sums to one embeds the assumption that the breadth of states is known, and CBDT seeks to relax this assumption.

Matsui (2000) provides a mapping from each CBDT instance to an Expected Utility Theory (EUT) instance which displays the same behavior. This implies that for any instance of CBDT (or CBSA), there exists an instance of EUT which displays the same choice behavior. Therefore, there can be no horse race between EUT and CBDT, as they cannot be empirically distinguished.[2] This does not mean that of the theories are redundant. On the contrary, Matsui says that "while one can embed a model based on one theory into a model based on the other theory, intuition which motivates the original model may be lost. It is intuition on phenomena rather than formal capability of description of certain situations that differentiates the two theories." This seems to be the case here. As we show in Section 4, 'similarity' is a concept familiar to psychologists, who have been able to empirically establish a functional form which appears to describe how humans conceive of similarity across a number of decision-making domains (Shepard 1987). Using this functional form and Matsui's transformation, it implies a specific form of non-diffuse Bayesian priors which should be preferred over diffuse priors, at least in this setting.

Agent-based Computational Economics and Artificial Intelligence. The central investigative tool of this paper is an artificial decision maker called the Case-based Software Agent or CBSA. CBSA is a software agent, which is "an encapsulated piece of software that includes data together with behavioral methods that act on these data," which interacts with its environment and/or other agents (Tesfatsion 2006). CBSA is thus part of Agent-based Computational Economics *(ACE)*. In fact, CBSA is the first example of agent-based computational economics making a direct contribution to decision theory.

Software agents are common in the field of artificial intelligence, which, like decision theory, also cites von Neumann as a founder. Indeed, 'decision theory' is a term used in artificial intelligence, and it refers to a different field which grew out of von Neumann's work.[3] The fields are largely distinct, with some notable exceptions. For example, Gilboa and Schmeidler (2000), in the artificial intelligence journal "IEEE

[1] CBU is defined slightly differently in the original paper: namely, CBU sums over the whole memory, assigning a utility of r_0 for those acts not preformed. However, since 'this action was not performed' can be assumed to have a value of $r_0 = 0$, it is irrelevant whether those cases are considered. Hence this formulation is equivalent.

[2] This result was noted in Gilboa and Schmeidler (1995).

[3] A field of the same name is also found in philosophy, and it has the same origin.

Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans," describe how CBDT is "closely related to (and partly inspired by) the theory of Case-Based Reasoning proposed by Riesbeck and Schank (1989) and Schank (1986)" which "is proposed as a better AI technology, and a more realistic descriptive theory of human reasoning than rule-based models (or systems)."

CBSA is written in the agent-based modeling platform NetLogo (Wilensky 1999).

3 Case-based Software Agent Description and Verification

In this section, we show that our computational implementation of CBDT, called the Case-based Software Agent or CBSA, correctly implements the representation specified in Gilboa and Schmeidler (1995). First we describe the primitives of the implementation. Then we describe the algorithms, which implement CBDT in this setting (as we argue in Section 3.2). We then express the formal relationship between CBSA and Expected Utility Theory using Matsui's (2000) transformation, which lays the groundwork for interpreting the implications of our empirical work using CBSA for EUT and Bayesian priors.

3.1 Primitives of the implementation

This implementation has three types of primitives: first, the primitives of CBDT; second, the CBDT representation; third, a decision environment.

The primitives and representation of CBDT define aspects of decision-making internal to the agent. The primitives of CBDT are: a finite set of actions $\mathcal{A}$, a finite set of problems $\mathcal{P}$, a set of results $\mathcal{R}= \mathbb{R}$, which includes $r_0= 0$ to be interpreted as "this action was not chosen," and the set of cases $\mathcal{C} = \mathcal{P} \times \mathcal{A} \times \mathcal{R}$. Moreover, a set $\mathcal{M} \subseteq \mathcal{C}$ is memory of this agent. From the CBDT representation, the agent is also endowed with a similarity function $s : \mathcal{P} \times \mathcal{P} \to \mathbb{R}_+$. The similarity value is interpreted as: the higher a similarity value $s(p, q)$, the more relevant is problem q to problem p in the mind of the decision-maker.

The decision environment defines those aspects of the series of problems the agent faces that are external to the agent. This environment both stands apart from the decision theory and is unknown to the agent.[4] We capture the decision environment with a function (algorithm) called the *problem-result map* or PRM. The PRM is the transition function of the environment. It takes as input the current problem $p \in \mathcal{P}$ the agent is facing, the action $a \in \mathcal{A}$ that the agent has chosen, and some vector $\theta \in \Theta$ of environmental characteristics. The PRM returns the outcome of these three inputs: namely, it returns a result $r \in \mathcal{R}$; the next problem $p' \in \mathcal{P}$ that the agent faces; and a potentially modified vector of environmental characteristics $\theta' \in \Theta$. I.e.:

$$PRM : \mathcal{P} \times \mathcal{A} \times \Theta \to \mathcal{R} \times \mathcal{P} \times \Theta$$

An example is in order. Consider the canonical Savage omelet problem (Savage 1954, pp. 13-15): the agent must choose to crack an egg in the main or secondary bowl, and the egg may be rotten or good. If the egg is good, the optimal behavior is to crack into the main bowl, and, if the egg is rotten, the optimal behavior is to crack into the secondary bowl. In the above language, the acts are Main or Secondary. A particular egg is a problem, and the similarity function may describe visual clues about eggs: shape and

[4]Indeed, all the CBDT decision-maker 'knows' about the problems she is facing is specified in the endowed similarity function above and her memory of cases at any given point in time.

color, say. Now consider a series of Savage omelet problems: suppose that the agent has a series of eggs delivered from a machine (a black box, as it were.) Θ would describe the internal state of this machine, which may be unknown to the agent: for example, the fraction of eggs it holds that are rotten. The agent faces a choice with regard to a particular egg, whether to crack into the main or secondary bowl. When the agent makes her choice, the machine provides the next egg. The PRM can be thought of as this black box: it provides an egg which is rotten or not (which embeds the result of each action) and maintains a θ which describes the internal state of the machine: namely, what eggs are rotten.

The PRM which corresponds to the classification learning problem from psychology is specified in Section 4.2.2.

3.2 Algorithms of the implementation

Input: problem p, memory $\mathcal{M}$.

1. For each $a \in \mathcal{A}$:
   (a) Construct $\mathcal{M}_a = \{(q, a, r) | \exists q \in \mathcal{P}, r \in \mathcal{R} \text{ s.t. } (q, a, r) \in M\}$
   (b) If $\mathcal{M}_a \neq \emptyset$,
   $$U_a = \sum_{(q,a,r) \in \mathcal{M}_a} s(p, q) r$$
   Else
   $$U_a = r_0 = 0.$$
2. Construct set $BEST$
   $$BEST = \left\{ a \in \mathcal{A} \middle| U_a = \max_{b \in \mathcal{A}} \{U_b\} \right\}$$
3. If $\#(BEST) = 1$ then let $a^\star$ be the sole entry in $BEST$. Else choose one element uniformly from the set $BEST$ and assign that to $a^\star$.

Output: Selected action $a^\star$

Figure 1: The Choice Algorithm

Figure 1 describes the choice algorithm which implements the core of CBDT. The agent faces a problem $p \in \mathcal{P}$ and has a memory $\mathcal{M} \subseteq \mathcal{C}$. For each action a that she has available to her, she consults her memory and collects those cases, $\mathcal{M}_a$, in which she performed this act. Then she uses this subset of her memory to construct an 'expected utility' of that act, called here U_a. This value is a similarity-weighted payoff: a sum across the cases in $\mathcal{M}_a$, the result times the similarity between the problem faced at that time and the current problem. The agent then chooses the action which corresponds to the maximum U. There is an additional step, left unspecified in the original CBDT: In the case of a tie, the agent randomizes uniformly over the acts which achieve this maximum.

Figure 2 describes a single choice problem faced by the agent. It imbeds a reference to the choice algorithm

Input: problem p, memory $\mathcal{M}$, characteristics θ.

1. Input p,$\mathcal{M}$ into choice algorithm (Figure 1). Receive output $a^\star$.
2. Let $(r, p', \theta') = PRM(p, a^\star, \theta)$.
3. Let $\mathcal{M}$' $= \mathcal{M} \cup \{(p, a^\star, r)\}$

Output: problem p', memory $\mathcal{M}'$, characteristics θ'.

Figure 2: A single choice problem.

described in Figure 1. While the choice algorithm describes CBDT in the sense that it describes decision making, the algorithm described in Figure 2 embeds the agent in an environment and explicitly references that environment, in the call to PRM. In step one, the agent selects a best act, $a^\star$. In step two, the action is performed, in the sense that the environment of the agent reacts to the agent's choice: the PRM takes the current problem p, the action selected by the agent $a^\star$, and the characteristics unobserved by the agent θ, and constructs a result r, a next problem p', and a next set of characteristics θ'. Finally, the agent's memory is augmented by the new case which was just encountered: that is, the case that was just experienced is added to the set $\mathcal{M}$. Note that the choice problem maps a problem, characteristic, memory vector to another vector in the same space, so it can be applied iteratively.

1. Initialize $\mathcal{M} = \emptyset$. Initialize θ. Initialize first problem p.
2. Evaluate single choice problem with input $(p, \mathcal{M}, \theta)$. Receive output $(p', \mathcal{M}', \theta')$.
3. Evaluate: Does (p', θ') achieve ending condition? If so, stop. Else, return to previous step with new vector $(p', \mathcal{M}', \theta')$

Figure 3: A complete series of choice problems.

Figure 3 describes a complete series of problems faced by an agent. Initially, the agent is assumed to have an empty memory, and some initial problem p.[5] Furthermore, it is assumed that there is some initial starting condition θ. Essentially, thereafter, the single choice problem is repeated iteratively. In step three p' and θ' are jointly evaluated against some ending condition, and if that ending condition is achieved, the algorithm halts.

3.3 Transforming CBDT into EUT

Every instance of CBDT can be transformed into an instance of EUT which exhibits identical behavior. This transformation, due to Matsui (2000), provides us with the instance of EUT described below. The purpose of this section is to use this transformation to illuminate implications of this paper for Bayesian models of learning. We follow Matsui's discussion, notation, and definitions from his Section 2.1.[6]

[5] It is simple to modify the algorithm such that the agent starts with some non-empty memory.

[6] For a more detailed explanation of the transformation and the proof that the transformation is behaviorally equivalent, we refer you to that Matsui's (2000) excellent paper.

Take as given an instance of CBDT $\{\mathcal{P}, \mathcal{A}, \mathcal{R}, s, u\}$, where u is a utility function $u : \mathcal{R} \to \mathbb{R}$.[7] Let $\left\{(\Omega, \mathcal{F}), \mathcal{A}, \bar{R}, f, \bar{u}, \mu\right\}$ be the corresponding instance of EUT by Matsui's transformation, where: Ω is a state space, $\mathcal{F}$ is a σ-algebra on Ω, $\mathcal{A}$ is the set of available actions,[8] $\bar{R}$ is a countable set of possible results, or, equivalently, pieces of information, $\bar{u} : \bar{R} \to \mathbb{R}$ is a utility function, and μ is a probability measure on $(\Omega, \mathcal{F})$.

We define a history h as a subset of cases $\mathcal{C}$. From this point of view, a memory is a history. Matsui's transformation is about behavioral equivalence after identical histories.

Matsui's transformation specifies the elements of $\left\{(\Omega, \mathcal{F}), \mathcal{A}, \bar{R}, f, \bar{u}, \mu\right\}$ as functions of the elements of $\{\mathcal{P}, \mathcal{A}, \mathcal{R}, s, u\}$ as follows:

$\bar{R} = \mathcal{R} \times \mathcal{P}$. This means an outcome for the EUT model is constructed from a pair of elements from the CBDT model. In particular, the EUT model encodes that a choice results in a 'result' from the set $\mathcal{R}$ and a 'next problem' from the set $\mathcal{P}$.

Define $f : \bigcup_{T=1}^{\infty} A^T \times \Omega \to \mathcal{R}$ is an outcome function where A^T is the cross product of A. This function closely mirrors the role of the PRM or Problem-Result Map defined above. It defines, for a series of actions and a particular state of the world, what result is awarded to the user. Define $f_h(a)$ as a description of Nature's response (a result $(r, p) \in \bar{R}$) from an agent's choice of a after some history h.

$\bar{u}(r, p) = u(r)$ for all $(r, p) \in \bar{R}$; that is, the utility function is essentially unchanged, and continues to all but technically apply only to elements of the set $\mathcal{R}$.

Now let $\bar{r}(p, a, h)$ be a possible result which occurs when action a is taken at encountering problem p after history h. Let ω be a typical member of the state space Ω. Let $\omega = (p_1, \omega_1, \omega_2, \ldots)$, where p_1 is the first period problem, and $\omega_t = \{\bar{r}(p, a, h)\}_{(p,a,h) \in \mathcal{P} \times \mathcal{A} \times H_t}$ for all $t = 1, 2, \ldots$—that is, ω_t is the set of all result/next-problem pairs that can occur, one for each problem-action-history combination for the problem faced at, and the action taken at, time t, after the history in H_t.

Then the state space is the set of all such elements $\omega = (p_1, \omega_1, \omega_2, \ldots)$. That is, $\Omega = \mathcal{P} \times \Pi_{T=1}^{\infty} \left(\bar{R}^{\mathcal{P} \times \mathcal{A} \times H_t}\right)$. That is, a state of the world is completely described by an initial problem and a complete description of all possible result/problem pairs after every possible sequence of play.

Given those definitions of objects, the main work of the proof is to construct a conditional probability measure μ_h for each possible history h which encodes a probability constructed from data (the history) and similarity. The probability is constructed in such a way as to yield the same maximization behavior. Note that this conditional probability is conditional on each possible history, so there is sufficient freedom to construct a probability measure which allows for identical revealed choice behavior between the EUT model and the CBDT model. The probability is constructed by converting the total similarity "weight" on an outcome to the revealed believed probability of that event occurring.[9] In this way, when the functional form of similarity is tested in this paper, it can be thought of as testing different forms of priors. Under this construction, the turn-by-turn maximization problems are equivalent. Note that, given this set of conditional probability measures, a total probability measure over Ω is easy to construct, and, by construction, satisfies Bayes' Rule. Therefore, the EUT decision-maker also follows Bayes' Rule, when the state space is defined according to Matsui's transformation.

[7] It is more convenient notationally to follow Matsui here, and induce a separation between the set of results and utility over those results.

[8] Note that the acts set $\mathcal{A}$ remains unchanged in Matsui's transformation.

[9] The specific mathematical formulation is not used in this paper, so we do not reproduce it here. Once again, we refer interested readers to Matsui (2000).

4 Psychology of Human Classification Learning

The classification problem is used regularly in psychology and cognitive science to study both natural and artificial agents (Pothos and Wills 2011). The classification learning problem presents a standard set of objects to the agent, who must sort the objects into bins. The correct sorts vary in complexity and are therefore harder or easier to learn. Psychologists have collected data on how humans solve these problems. These data include variables such as probability of errors over time. Psychologists have also developed a number of Models of Classification Learning (MCLs), which, like CBSA, are mathematical models that are implemented as software programs to generate simulated choice data. Psychologists compare the fit of MCLs to human data to evaluate the MCLs' explanatory power. We evaluate CBSA in the same way to establish our central results.

Game theorists might describe the classification problem as: a choice problem of sorting a series of vectors into bins, where Nature provides a true mapping of vectors to bins, and the agents' utility is determined by her accuracy in following Nature's mapping. The classification problem is familiar to decision theory. Savage's famous omelette problem, in which he sorts rotten eggs from fresh ones (Savage 1954, pp. 13-15), can be considered a classification problem.

Models of classification learning are related to decision theories but they are not identical. On one hand, DTs and MCLs are both mathematical models designed to represent human decision making. On the other hand, decision theories and MCLs have different sets of primitives. Expected Utility Theory has the following primitives: a set of acts or actions which represent the choices available to the decision maker; a set of outcomes or payoffs or results; and a set of states or problems which provide a mapping from actions to outcomes. MCLs have the a different set of primitives: a set of objects; a set of categories; and a mapping from objects to categories which represents the correct categorization of these objects.

4.1 The SHJ Classification Experiment

Shepard, Hovland, and Jenkins (1961) performed a canonical laboratory experiment which established some empirical facts about human classification learning. In this study, Shepard, Hovland, and Jenkins (1961) (hereafter SHJ) introduced a set of objects to be sorted: eight elementary objects with three binary dimensions of shape (square or triangle), color (dark or light), size (large or small). These eight objects can be thought of as three-digit binary strings. Figure 4 is an illustration of these objects as they are typically represented in the psychology literature: each binary string is placed on a vertex of the unit cube. SHJ

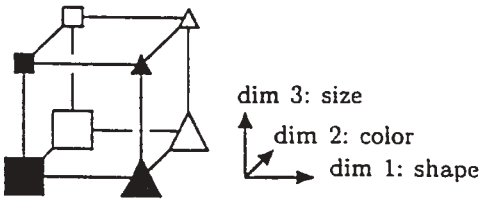

Figure 4: The eight elementary objects introduced by SHJ.
(Image from N94)

also introduced a now standard set of mappings or classifications $T = \{I, II, \ldots, VI\}$. SHJ named the six

b digits			Category, by Mapping Type					
1	**2**	**3**	*I*	*II*	*III*	*IV*	*V*	*VI*
0	0	0	O	O	O	O	O	O
0	0	1	O	O	O	O	O	R
0	1	0	O	R	O	O	O	R
0	1	1	O	R	R	R	R	O
1	0	0	R	R	R	O	R	R
1	0	1	R	R	O	R	R	O
1	1	0	R	O	R	R	R	O
1	1	1	R	O	R	R	O	R

Table 1: The complete description of the six mappings Type I-VI of three-digit binary strings to categories.

mappings "Problems of of Type I through VI." It is the relative and absolute performance of sorting under each of these six mappings that we measure here. Figure 5 shows the six mappings where category one is represented by ovals and category two is represented by rectangles. Table 1 displays the same six mappings

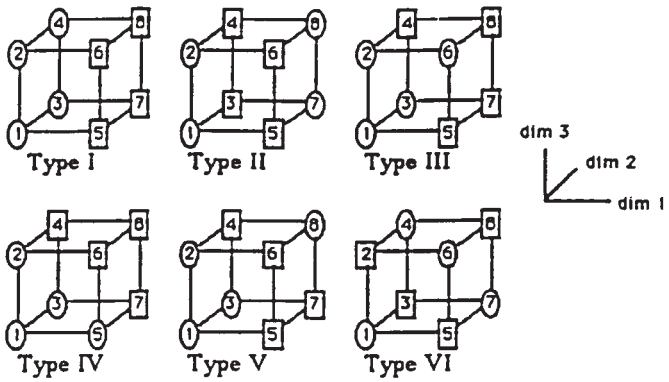

Figure 5: The six classifications of the elementary objects into categories.
(Image from N94)

as a table. The eight objects as three-digit binary strings are listed in the first three columns. The second group of columns represents the category of oval O or rectangle R assigned to each binary string under each mapping.

Consider the relatively simple Type I mapping, which can be seen as the first cube in Figure 5 or the fourth column in Table 1. In this mapping only the first dimension, *shape*, is required to sort the objects correctly. In Figure 5, this can be seen by the left face of the cube being marked with ovals and the right face being marked with rectangles. Correspondingly, in Table 1, it can be seen as the first four lines being ovals O and the second four lines being rectangles R.

Now consider the relatively complicated Type VI mapping, which can be seen as the sixth cube in Figure 5 or the last column in Table 1. In contrast with mapping Type I, in mapping Type IV all three dimensions are required to sort the objects correctly. This can be seem most plainly in the sixth cube in Figure 5.

The intuition that Type VI is more complicated than Type I is reflected in human performance. In

repeated trials, humans learn Type I problems much faster than Type VI. In these trials, images of these eight objects were shown to people in a laboratory setting in a random order, and, after each category guess, the subjects were told whether or not they were correct. MCLs are put through the same process and given the same feedback. Whether MCLs achieve the same ordering over problems is an important metric of success in explaining human data in the psychology literature, and it is a metric we use here (see Section 5).

4.2 Classification Learning and Case-based Decision Theory

The purpose of this section is to interpret the classification learning problem as understood and tested by psychologists into the Case-based Decision Theory/CBSA framework, in order to 'run' CBDT on the classification problem in a way to generate simulated data comparable to human data and those simulated data generated by other MCLs. We proceed in three steps. First, we relate some fundamental concepts in the psychology of classification learning to decision theory. Second, we formally define the the problem-result map (PRM), the choice setting facing CBSA, which we construct to be consistent with the SHJ series of classification learning experiments. Third, we choose a functional form of similarity based on empirical work in psychology. (We test this functional form of similarity against alternatives.)

4.2.1 Related concepts in Psychology and Decision Theory

The concept of a *problem* in CBDT can be mapped to the concept of a *stimulus* in psychology. For example, in describing the classification learning experiment, a psychologist would refer to an object shown to a decision-maker as a stimulus.

The concept of the *problem set* $\mathcal{P}$ can be mapped to the psychological concept of a *psychological space*. However, psychologists impose more structure on a psychological space than does CBDT on a problem set: A psychological space is assumed to be an n-dimensional space, in which each dimension represents some characteristic of the stimuli. By contrast, the problem space in CBDT need not have any particular structure.

In SHJ, since the objects have three dimensions, the psychological space is a subset of $\mathbb{R}^3$. Since the dimensions are binary, the unit cube is the psychological space of the classification problem. This is why psychologists chose to represent the eight three-digit binary strings to the vertices of a cube (as seen in Figure 4). It also implies that the natural sense of distance is geometric distance: either L2 (Euclidean) or L1 (Manhattan) distance. This point becomes relevant when we discuss similarity functions, some of which incorporate geometric distance.

Psychologists posit that humans make guesses about the characteristics of new stimuli from old stimuli, a process which they call 'generalization.' To *generalize* from a data set in memory to a new problem is to make predictions about the result of actions undertaken with this new problem. Psychologists studying generalization seek empirical regularities of generalization and even a "universal law of generalization." (Shepard 1987). Case-based Decision Theory describes one way humans might generalize.

4.2.2 The Problem-Result Map (PRM) implied by the Classification Learning Experiments

The Problem-Result Map, or PRM (see Section 3), takes as input {an action $a \in \mathcal{A}$, a problem $p \in \mathcal{P}$, and a set of environmental characteristics $\theta \in \Theta$,} and delivers {a result in $r \in \mathcal{R}$, a 'next problem' p' the agent is to face, and a potentially modified vector of environmental characteristics $\theta' \in \Theta$}. In this section, we define

the primitives of CBDT needed to describe the classification problem introduced by SHJ and then define the PRM.

We define the primitives as follows: Let $\mathcal{P}$ be the complete set of three-digit binary strings (therefore $\#(\mathcal{P}) = 2^3 = 8$). Let $\mathcal{A} = \{a_1 = O, a_2 = R\}$ be the set of categories. There is a correct mapping, given by Nature, from strings $p \in \mathcal{P}$ to categories $a \in \mathcal{A}$ called $F : \mathcal{P} \to \mathcal{A}$. Each mapping F is called a *classification*, and some Fs may be harder to learn than others. We use F_I through F_{VI} to evaluate CBSA; these mappings correspond to the SHJ "Problems of Type I through IV" described above.

Briefly, the choice problem proceeds as follows: Nature selects a string $p \in \mathcal{P}$ and presents p to the agent. The agent observes p, and announces a category guess $a^\star \in \mathcal{A}$. His payoff is determined by whether his guess is correct, and then Nature presents the agent with a new problem p'.

The environmental characteristic $\theta \in \Theta$ describes those aspects of the environment that are changing over time. In the classification problem, the only aspect of the environment that is changing over time is the randomization device: i.e. how Nature determines the next binary string $p \in \mathcal{P}$ that will be encountered. We assume that Nature selects $p \in \mathcal{P}$ according to some distribution; we let Θ is the set of all distributions over $\mathcal{P}$, so at any given point in time, the next problem will be selected according to the current distribution $\theta \in \Theta$.

The PRM in Figure 6 describes how the environment responds to the agent choice.

In Step 1, the payoff (result r) is determined in the following way: if $a^\star = F(p)$, then the agent's category guess matches the true category for that string, and his answer is 'correct,' and the agent receives a utility payoff of $r_0 = 0$. If $a^\star \neq F(p)$, then the guess is 'incorrect' and the agent receives a payoff of -1.

In Step 2, the next problem p' is selected. As mentioned above, the environmental characteristic θ provides the distribution which is used to select p'.

In Step 3, the next distribution is specified. We are not free to choose θ'; instead, the series of θs must correspond to how the problems/strings were selected in the actual SHJ experiments. The new environmental characteristic $\theta' \in \Theta$ is given by: if all the elements of $\mathcal{P}$ have been selected, then θ' is set to a uniform distribution over $\mathcal{P}$.[10] Otherwise, set to zero the probability of the problem just selected, p', and reweight the remaining probabilities according to Bayes' Rule. This follows the procedure of SHJ, in which the objects were selected uniformly but without replacement.[11]

4.2.3 The Functional Form of the Similarity suggested by Psychology

Psychology has done a good deal of research about the functional form of similarity common to humans. Shepard (1987), in the journal *Science*, finds that "[e]mpirical results and theoretical derivations point toward two pervasive regularities of generalization." First, he finds that similarity "approximates an exponential decay function of distance in psychological space." Second, he finds that, "to the degree that the spreads of consequential stimuli along orthogonal dimensions of that space tend to be correlated or uncorrelated,

[10] Any distribution over $\mathcal{P}$ could be substituted here, if desired. The uniform distribution is chosen to be consistent with the human experiments.

[11] This description is slightly simplified. For the interested reader, we describe the exact randomization process: The first sixteen trials consist of: the first eight trials consist of all elements of $\mathcal{P}$ in a random order, and the second eight trials are again, all elements of $\mathcal{P}$ in a random order. In the second set of sixteen trials, each element of set of $\mathcal{P}$ is presented twice in a random order. All subsequent sets of sixteen trials follow the same randomization procedure as the second set, until the experiment concludes. CBSA always converges in the first sixteen trials, however, so this distinction has no impact on the results presented here.

Input: act a, problem p, characteristics θ.

1. Set result r.

$$\text{Let } r = \begin{cases} r_0 = 0, & \text{if } a = F(p) \\ -1, & \text{if } a \neq F(p) \end{cases}$$

2. Set next problem p', by drawing p' from $\mathcal{P}$ according to the distribution θ.

3. Set next characteristic θ'.

   If θ is a degenerate distribution, then

$$\text{Let } \theta'(q) = \frac{1}{\#(\mathcal{P})} \qquad \forall q \in \mathcal{P}$$

   Else

$$\text{Let } \theta'(q) = \frac{\theta(q) \cdot 1_{q \neq p}}{\sum_{q' \in \mathcal{P}} \left[\theta(q') \cdot 1_{q' \neq p}\right]} \qquad \forall q \in \mathcal{P}$$

   Where 1_x is the indicator function.

Output: result r, problem p', characteristics θ'.

Figure 6: The Classification Learning PRM

psychological distances in that space approximate the Euclidean or non-Euclidean metrics associated, respectively, with the L_2- and L_1-norms of that space."

In the classification problem, the distribution, dimensionality, and psychological space of the stimuli are carefully controlled. In this case, Shepard's results have some remarkably specific implications about the functional form of similarity. Shepard's first result is that similarity ought to be measured by the inverse exponential of psychological distance. Since the stimuli are randomly presented to the agent in a manner that makes them uncorrelated across dimensions, Shepard's second result is that a Euclidian distance between stimuli is an appropriate choice for the psychological distance in this case. Applying these results, our primary similarity function, which we simply call s, is:

$$\begin{aligned} s(p,q) &= \frac{1}{e^{d(p,q)}} \\ \text{where } p,q &\in \{0,1\}^3 \\ \text{and } d(p,q) &= \sqrt{\sum_{i=1}^{3}\left[(p_i - q_i)^2\right]} \\ & \text{and } p_i \text{ refers to the } i^{\text{th}} \text{ element of } p \end{aligned} \tag{1}$$

(Note: $d(p,q)$ is standard Euclidean or $L2$ distance.) This is the similarity function we use to generate the primary results below. We test this form of similarity against others. These results are described in Section 5.2.

4.3 The stylized facts of SHJ and related experiments

Here we describe the core phenomenology or stylized facts of human classification learning that have been empirically established with the SHJ problem set. The original findings of Shepard, Hovland, and Jenkins (1961) have been replicated using contemporary research methodology as a benchmark for comparison of formal models (Nosofsky, Gluck, Palmeri, McKinley, and Glauthier 1994).

$$I < II < III, IV, V < VI \tag{2}$$

The key findings are the ease of Type I learning, the difficulty of Type VI learning, approximately equal difficulty of Types III, IV, V, and a Type II advantage relative to Types III, IV, V. These results have been replicated repeatedly except for the Type II advantage which appears to depend on particular experimental conditions (Kurtz, Romero, Stanton, and Morris 2011). Leading theoretical accounts suggest that this pattern of performance is attributable to some type of psychological mechanism that allows human learners to focus their attention on, verbalize, or abstract rules or regularities about particular dimensions (characteristics) of the stimuli.

Fortunately for present purposes, Nosofsky and Palmeri (1996) studied human learning of the same six category structures using a different set of stimuli than the eight objects described above based on the dimensions of shape, color, and size. The canonical stimuli are based on easily separable dimensions, that is, a human viewer can easily identify and characterize each item in terms of its dimension values, as opposed to integral dimensions, which are difficult for a human viewer to distinguish. Shepard and Chang (1963)

	Total	I	IV	III	V	II	VI
MEAN	6.42	3.69	5.51	6.50	7.13	7.65	8.04
STD	1.82	1.05	1.40	1.24	1.00	1.14	0.20
MIN	2	2	2	4	6	4	8
MAX	10	6	7	8	10	8	9
N	12000	2000	2000	2000	2000	2000	2000

Table 2: CBSA's # Errors For Each Problem Type.
All means pass pairwise T-tests at the 0.01% *level.*

predicted that basic principles of stimulus generalization could account for human classification learning performance with integral-dimension stimuli, but that a focusing mechanism would be needed to explain classification learning with separable-dimension stimuli. Nosofsky and Palmeri (1996) tested this prediction on the SHJ problem set using integral dimensions of hue, saturation, and brightness, which are not readily picked out for separate analysis by the learner. They found a different ordering of difficulty, one consistent with pure stimulus generalization theory, confirming Shepard and Chang's hypothesis. The critical differences in the ordering are that Type II is the second-most difficult to learn and reliable differences are found between Types III, IV, and V. In addition, the learning is slower in general.

$$I < IV < III < V < II < VI \tag{3}$$

In its pure form, CBSA does not have a focusing mechanism, because the similarity is assumed to be fixed. Therefore, the appropriate human data with which to test CBSA is the integral-stimuli benchmark (Nosofsky and Palmeri 1996). As reported below, CBSA produces an impressive fit to these data. In future work, we will implement and test an extension to CBSA that can endogenously change its similarity function to increase the weight on certain dimensions and reduce the weight on others. It is expected that this version of CBSA will successfully predict the original SHJ ordering.

5 Experimental Results

Table 2 contains the summary statistics of the number of errors CBSA committed with each problem type, using the Shepard similarity function (inverse exponential Euclidean distance.) Two thousand runs of each problem type were performed, where a 'run' is defined as 100 iterations of the PRM. Agents converge to behavior making zero mistakes well before the one-hundredth choice, so the 100 iterations limit is more than sufficient. Runs differ because of two stochastic elements: (a) as specified in the PRM, order in which the agent faces problems is random, and (b) as specified in the CBSA choice algorithm, when CBSA is indifferent between two choices of action, the agent randomizes between them. There are no other sources of randomness.[12]

All pairwise t-test comparisons among number of errors for different problem types are significantly different at the 0.01% level. Figure 7 contains the raw data from human trials from NP96, to be compared to the data presented in Figure 8. Only the probability of error during the first block is presented, because

[12]Since this is a computer simulation, 'random' in fact means pseudorandom.

the probability of error in all subsequent blocks was zero.

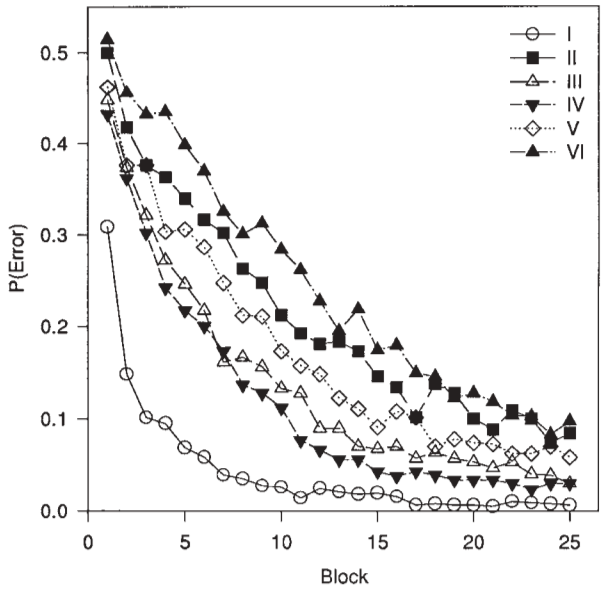

Figure 7: Average Probabilities of errors in each problem in each block of 16 trials. (Human data. Source: NP96.)

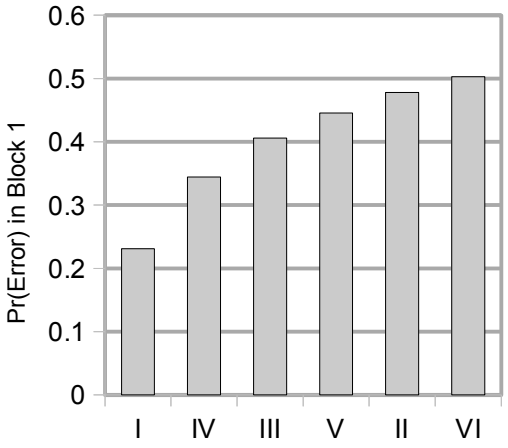

Figure 8: Average Probabilities of errors for each Problem Type in First Block. (Simulated data: CBSA.)

5.1 Main results

Relative Difficulty. The order of means in Table 2, from easiest to hardest, is Type I < Type IV < Type III < Type V < Type II < Type VI. This matches the empirical order found in NP96, in which humans were not able to easily distinguish between the dimensions. The relative size of these gaps–that the gap between Type 1 and the next is about twice that of all subsequent gaps, which are of similar magnitude–also matches human data. See Figures 7 and 8, where this pattern can be seen.

Overall Speed. Humans make mistakes well into their tenth "block" of sixteen trials. CBSA never makes more than ten mistakes, and never after the first "block." CBSA is therefore orders of magnitude faster than humans on the same problem. No statistical test is necessary, because the support of the distributions do not overlap. This can be taken as strong evidence that a more realistic decision theory would incorporate some kind of error, either at the point of observing objects, at the point of action selection (a 'trembling hand'), at the point of receiving a result, or at the point of either adding the case to memory or consulting memory.

One kind of 'trembling hand'-style error that may be useful to incorporate is Luce's choice rule (Luce 1959, Luce 1977). Luce's choice rule supposes that instead of choosing the preferred action with certainty, the agent chooses the preferred action with a likelihood that is increasing in the preference. For example, the Luce choice rule in this setting might suggest that

$$\text{Probability of selecting action } a = \frac{CBU(a)}{\sum_{\alpha \in \mathcal{A}} CBU(\alpha)}$$

Problem Type	Avg Num Errors
I	3.69
IV	5.51
III	6.50
V	7.13
II	7.65
VI	8.04

(a) Sim. s

Problem Type	Avg Num Errors
I	3.67
IV	5.45
III	6.45
V	7.13
II	7.6
VI	8.03

(b) Sim. s'

Problem Type	Avg Num Errors
I	3.69
IV	6.14
III	6.81
V	7.95
II	9.10
VI	9.54

(c) Sim. s''

Problem Type	Avg Num Errors
I	8
II	8
III	8
IV	8
V	8
VI	8

(d) Sim. $s_=$

Problem Type	Avg Num Errors
IV	2.76
VI	5.11
I	5.88
V	6.25
III	6.37
II	8

(e) Sim. s_+

Problem Type	Avg Num Errors
I	37.28
IV	37.86
III	38.12
V	38.32
II	38.62
VI	40.31

(f) Sim. s_{avg}

Figure 9: Relative difficulty of Problems Type I-VI for a selection of similarity functions.

where CBU_a is the calculated Case-based Utility associated with action a. Luce's choice rule has received attention in economics in empirical study of consumer choice, but has not received much attention within decision theory.

5.2 Functional Form of Similarity

The functional form of similarity is critical to get the difficulty ordering to match human data from NP96. Shepard (1987) recommends the inverse exponential of Euclidean distance (i.e. $e^{-d(p,q)}$) for matching human notions of similarity: so the closer in Euclidean distance, the more similar, at a maximum similarity of $1 = e^0$. (As previous, $d(p, q)$ is defined as standard Euclidean vector distance.)

The exponential functional form of distance appears not to be critical. We test against two alternatives to the inverse exponential function s recommended by Shepard. The two alternatives considered here are of a simple inverse and inverse log, which we call s' and s'' respectively:

$$s'(p, q) = \frac{1}{d(p, q) + 1} \tag{4}$$

$$s''(p, q) = \frac{1}{\ln(d(p, q) + 1) + 1} \tag{5}$$

(Ones are added to avoid division-by-zero and log-of-zero problems.) In these results, there is no difference in the relative difficulty ranking. This likely results from the fact that, since there are only eight objects to be judged to be similar, the distinctions between these functions are too fine for it to have an impact. (It may have an impact once error is introduced, in matching the shape of learning curves.) See Figures 9(a) versus Figure 9(b) and Figure 9(c) to see the results of s versus s' and s''.

On the other hand, geometric distance, either Euclidean (i.e. the $L2$ norm) or the $L1$ norm, seems to

be critical. Essentially, the underlying distance metric is the agents' mental model of what objects tend to have the same properties as others. This dictates which objects the agent extrapolates (or generalizes) to first. To understand the impact, it is easiest to consider the similarity function $s_=$, in which all objects are similar to themselves but all other objects are equally dissimilar.

$$s_=(p,q) = \begin{cases} 1 & \text{if } p = q \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

An agent does not extrapolate from any object to any other object. For this agent, problems of type I through VI are all equally difficult: This agent makes exactly 8 mistakes in each run, which corresponds to learning about each object once. See Figure 9(d).

It is also possible to choose similarity functions which invert important components of the ordering. For example, consider similarity function s_+:

$$s_+(p,q) = \frac{1}{2\left|\sum_{i=1}^{3} p_i - \sum_{j=1}^{3} q_j\right| + d(p,q)} \tag{7}$$

where d is the standard Euclidean distance function. This is a similarity function that declares vectors are more similar when their entries sum to the same number (hence named s_+); i.e. it judges $(1,1,0)$ as fairly similar to $(0,1,1)$, because their entries both sum to 2. See Figure 9(e) for the impact on relative ordering. Comparing to Figure 9(a), one can see that the organizing principle encoded in the similarity function s_+ assists the agent in solving problem Type VI, as the agent makes almost three fewer errors on this problem, but makes solving problem Type I more difficult: the agent makes almost two and a half more errors. This inverts the ordering that holds for humans across all versions of this experiment: for humans, problem Type I is always easier than problem Type VI, but not under similarity function s_+. This implies that s provides a better fit for human data than does s_+.

Average similarity is an alternative similarity function proposed in Gilboa and Schmeidler (1995). It has the following functional form:

$$s_{avg}(p,q) = \frac{s(p,q)}{\sum_{(q',a,r)\in\mathcal{M}} s(p,q')} \tag{8}$$

As Matsui (2000) points out, when an instance of CBDT is transformed into EUT, one finds the implied belief distribution over states after a history h, called μ_h, goes up with similarity and data accumulation. Average similarity removes one of those aspects: it removes the data accumulation aspect. As we see in Figure 9(f), this dramatically increases the failure rate of the agent, while not changing the underlying order (i.e. the ordering agrees with similarity function s.) The reason this occurs is, in about seventy percent of cases, the average similarity agent immediately settles on one preferred category which it repeats forever. It never converges to correct behavior, which is inconsistent with human behavior on this problem. It appears that removing the data accumulation aspect of similarity strongly turns agent behavior away from human-like behavior. Agent-based modelers discuss a trade-off between 'exploration' and 'exploitation' with regards to difficult problems. The value of exploring the solution space is the possibility of finding a better solution than the current best solution, and the cost is not exploiting the current best solution. Average similarity

appears to over-emphasize the value of exploitation relative to exploration. The agent with average similarity apparently quickly decides that the correct model of the world is very simple: one category is better than the other category and some level of errors are unavoidable.

5.3 Comparison to Cognitive Models

A leading formal model of classification learning known as ALCOVE (Kruschke 1992) has been successfully fit to the classic SHJ ordering and the integral-stimuli version (Nosofsky and Palmeri 1996). ALCOVE is an adaptive network model (artificial neural network) that instantiates the exemplar theory of categorization (Medin and Schaffer 1978, Nosofsky 1986). The exemplar view states that the psychological representation of a category is the stored examples themselves (i.e., no abstractions are formed) and the process of classification is based on the similarity between each stored example and the stimulus (in the sense of similarity used in this paper). Therefore, if a stimulus is highly similar to one or more exemplars that are associated with category A, then the presentation of that stimuli will strongly activate category A. The trial-by-trial learning process in ALCOVE consists of storing each new example that is experienced and using error-driven learning to optimize: 1) the associative mapping between examples and category labels; and 2) attention weights for each dimension to optimize the functional form of similarity. ALCOVE employs four free parameters for the rate of learning of association weights, the rate of learning of attention weights, the degree of specificity with which exemplars are activated by similar stimuli, and a bias on the mapping between category label activation and actual response behavior. When these parameter settings are optimized to match human data, high attentional learning and high specificity allow ALCOVE to best fit the classic SHJ ordering, while low (zero) attentional learning and low specificity produce a close fit for integral-stimuli SHJ (Nosofsky, Gluck, Palmeri, McKinley, and Glauthier 1994, Nosofsky and Palmeri 1996). In addition, the association learning rate and the response mapping constant allow ALCOVE to be calibrated to match overall learning performance.

This comparison to existing psychological models (ALCOVE, in particular) is useful for several purposes. First, it helps to contextualize the performance of CBSA. ALCOVE is one of only a handful of computational models that provide a good fit to the classic SHJ ordering – and it is the only one that has been shown to also fit integral-stimuli SHJ. Second, we note the theoretical correspondence between CBSA and ALCOVE (moreover to the exemplar view in general) in terms of the core explanatory principle of computing outcomes based on similarity to stored cases. Along these lines, we point out an intriguing difference between the approaches: while ALCOVE stores every case that is presented, in this implementation, CBSA only stores cases that lead to mistakes.[13] This is a compelling distinction to explore in further investigations. A successful model of category learning called SUSTAIN (Love, Medin, and Gureckis 2004) has much in common with ALCOVE, but employs the principle of storing an example when 'surprised' by its category label. Third, the design features of ALCOVE suggest clear directions for enhancing CBSA and improving the quality and range of its explanatory power by implementing extensions corresponding to the free parameters in ALCOVE.

[13]This arises from a correct guess being awarded $r_0 = 0$, the same outcome which is assigned to unknowns.

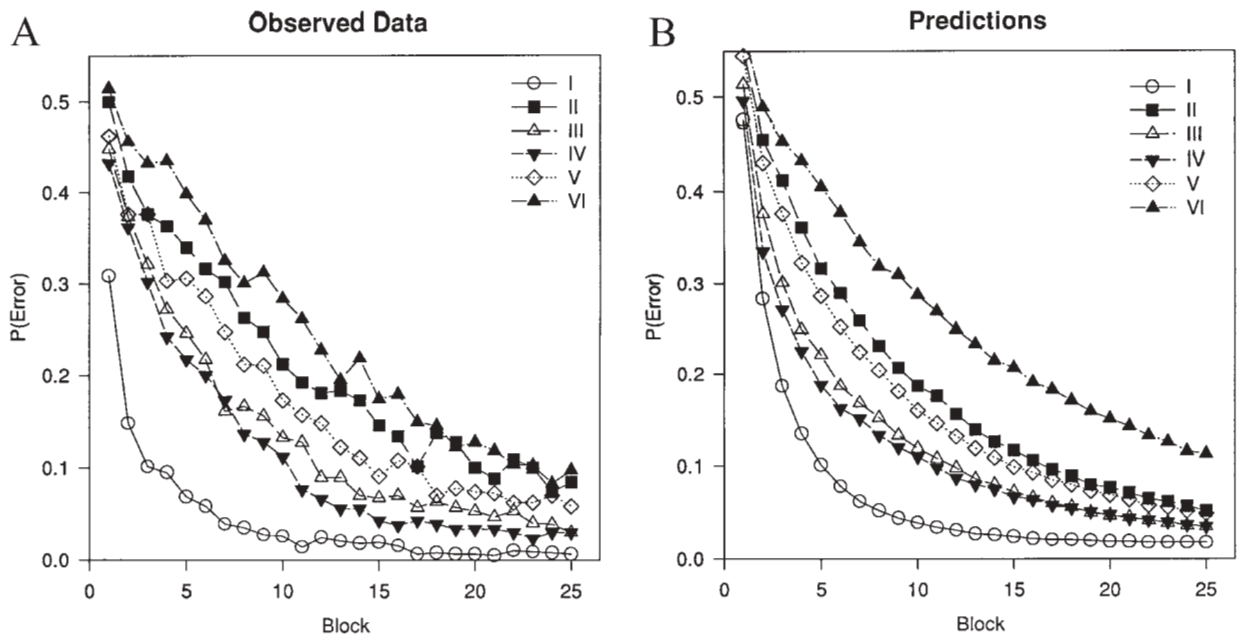

Figure 10: Human versus simulated ALCOVE results, N94

6 Conclusion

We present a computational implementation of Case-based Decision Theory (Gilboa and Schmeidler 1995) that can generate choice data consistent with an instance of CBDT for an arbitrary choice problem. The implementation is called the Case-based Software Agent or CBSA. We use CBSA to test the performance on a benchmark set of problems from the psychological literature on human classification learning exemplified by Shepard, Hovland, and Jenkins (1961) and Nosofsky, Gluck, Palmeri, McKinley, and Glauthier (1994). We find the choice behavior of CBSA (and therefore Case-based Decision Theory) appears to be psychologically valid with regards to the relative difficulty of these problems in a manner consistent with the human choice data in the study by Nosofsky and Palmeri (1996). On the other hand, we show that CBSA is psychologically invalid with regards to the speed of solving these problems. (CBSA is unrealistically fast.)

These results suggest many avenues of future research.

First, as suggested above, there is a strong reason to suspect that implementation of empirical similarity (Gilboa, Lieberman, and Schmeidler 2006), or some other means of making the similarity function endogenous to the agent's experience, will bring CBSA (and CBDT) closer to the classic result found by Shepard et. al.

Second, the result that CBSA is much faster at solving this problem than people suggests that some kind of error should be introduced into CBDT, if one is interested in replicating human choice.

Third, since CBSA is a software agent designed for an arbitrary choice problem, it can be used in any agent-based model where there is learning and choice under uncertainty. Moreover, since CBSA is an implementation of CBDT, it has a firm decision-theoretic foundation. This means that CBSA should be appealing to economists who would like to run an agent-based model with sophisticated agents that are both

economically and psychologically valid.[14]

Fourth, the decision theory characterization of the classification problem presented in this paper suggests that models of classification learning from cognitive science could be applied to a wider range of choice problems. This suggests that the decision theory approach may allow existing cognitive science models of classification learning to have wider applicability and be measured directly against other psychological data. As CBSA is already a decision theory, it, too, can also be measured against these same experimental data.

[14]CBSA is available for this purpose. Please contact the author.

References

Fudenberg, D., and D. Levine (2006): "A Dual-Self Model of Impulse Control," *The American Economic Review*, 96(5), 1449–1476.

Gilboa, I., O. Lieberman, and D. Schmeidler (2006): "Empirical Similarity," *The Review of Economics and Statistics*, 88(3), 433–444.

Gilboa, I., and D. Schmeidler (1995): "Case-Based Decision Theory," *The Quarterly Journal of Economics*, 110(3), 605–39.

——— (2000): "Case-Based Knowledge and Induction," *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, 30(2), 85–95.

Kruschke, J. (1992): "ALCOVE: an Exemplar-Based Connectionist Model of Category Learning.," *Psychological Review*, 99(1), 22.

Kurtz, K. J., K. L. J. Romero, R. D. Stanton, and S. N. Morris (2011): "Human Learning of Elemental Category Structures: Revising the Classic Result of Shepard, Hovland, and Jenkins (1961)," Submitted.

Laibson, D. (1997): "Golden Eggs and Hyperbolic Discounting*," *The Quarterly Journal of Economics*, 112(2), 443–477.

Love, B., D. Medin, and T. Gureckis (2004): "SUSTAIN: A Network Model of Category Learning," *Psychological Review*, 111(2), 309.

Luce, R. (1959): *Individual Choice Behavior*. John Wiley.

——— (1977): "The Choice Axiom After Twenty Years," *Journal of Mathematical Psychology*, 15(3), 215–233.

Matsui, A. (2000): "Expected Utility and Case-Based Reasoning," *Mathematical Social Sciences*, 39(1), 1–12.

Medin, D., and M. Schaffer (1978): "Context Theory of Classification Learning," *Psychological Review*, 85, 207–238.

Nosofsky, R. (1986): "Attention, similarity, and the identification–categorization relationship," *Journal of Experimental Psychology: General*, 115(1), 39–57.

Nosofsky, R., M. Gluck, T. Palmeri, S. McKinley, and P. Glauthier (1994): "Comparing Models of Rule-Based Classification Learning: a Replication and Extension of Shepard, Hovland, and Jenkins (1961)," *Memory and Cognition*, 22, 352–352.

Nosofsky, R., and T. Palmeri (1996): "Learning to Classify Integral-Dimension Stimuli," *Psychonomic Bulletin and Review*, 3, 222–226.

Pothos, E., and A. Wills (2011): *Formal Approaches in Categorization*. Cambridge University Press.

Riesbeck, C. K., and R. C. Schank (1989): *Inside Case-Based Reasoning*. Lawrence Erlbaum Assoc., Hillsdale, NJ.

Savage, L. J. (1954): *The Foundations of Statistics*. Wiley.

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

Shepard, R. (1987): "Toward a Universal Law of Generalization for Psychological Science," *Science*, 237(4820), 1317.

Shepard, R., and J. Chang (1963): "Stimulus Generalization in the Learning of Classifications," *Journal of Experimental Psychology*, 65(1), 94–102.

Shepard, R., C. Hovland, and H. Jenkins (1961): "Learning and memorization of classifications," *Psychological Monographs*, 75, 1–41.

Tesfatsion, L. (2006): *Handbook of Computational Economics*.vol. 2, chap. 16. Elsevier B.V.

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

Wilensky, U. (1999): *NetLogo*. http://ccl.northwestern.edu/netlogo/ Center for Connected Learning and Computer-Based Modeling, Northwestern University. Evanston, IL.