A value-relativistic decision theory predicts known biases in human preferences

Nisheeth Srivastava (nsriva@cs.umn.edu)
Department of Computer Science, 200 Union Street SE
Minneapolis, MN 55455 USA

Paul R Schrater (schrater@umn.edu)
Department of Psychology, 75 East River Road
Minneapolis, MN 55455 USA

Abstract

Traditional models of decision-making assume the existence of an external frame of reference for measuring the value of possible outcomes. We show that this fundamental assumption prevents classical decision models from predicting realistic decision-making behavior. Making an alternative relativistic assumption about the nature of reward leads us to formalize a rational agent as one which minimizes its internal decision-computational costs while retaining satisfactorily predictive models of its external environment. In computational evaluation, our model replicates previously unexplained 'irrational' behavior of human subjects.

Keywords: Cognitive models; neural coding; decision theory; behavioral biases

Introduction

Economists have used the terms 'animal spirits' and 'cognitive biases' (Akerlof & Shiller, 2009) to explain the persistence of behaviors incompatible with accepted definitions of rationality. Consistent observation of such 'predictably irrational' (Ariely, 2009) behavior in both controlled and uncontrolled settings has reduced confidence in the neo-classical view of human decision-making as a rational enterprise.

Contemporary explanations for these biases typically draw on evolutionary arguments (Gigerenzer & Goldstein, 1996), especially the idea of that deviations from traditional rationality represent specialized adaptations to the social and environmental conditions of mankind's early history. Thus, while on one hand, rational models of decision-making lie discredited through their inability to explain the existence of cognitive biases, the prominent alternative approaches create heuristic-based theories that are limited in their ability to produce generalizable causal explanations for decision-making processes. The absence of a realistic, principled theory of human decision-making is deeply problematic, since models of decision-making are central to the formulation of immensely consequential social and economic policies.

Given the continued failure of variants of existing theories to offer compelling explanations for how biological agents make decisions, we believe that it is necessary to re-evaluate the foundational assumptions in these theories and find plausible alternatives. How does an agent represent its possible options in an environment? How does it assign value to various options? On what basis does it select between these options? The canonical responses to these questions, obtained through a history of research stretching from J S Mill (Mill, 1874) to von Neumann (Neumann & Morgenstern, 1953), are that option possibilities are represented as environmental states, value is assigned to these options in the form of numerical reward, and the goal of the agent is assumed to be the maximization of its long-term cumulative *reward*. These assumptions are foundational to both *homo economicus* (Persky, 1995) models of economic choice and reinforcement learning (Barto & Sutton, 1998). In this paper, we question each of these three dogmas and replace them with alternatives obtained from a relativistic view of both how agents evaluate possible outcomes and how they evaluate their own *existence* as predictive agents.

At a fundamental level, the assignment of reward values assumes the existence of some fixed reference against which the value of its actions can be measured. This simple assumption causes us to formulate rewards as absolute numeric quantities, with a "no reward" state as the origin. Furthermore, this view renders us unable to consider any meaningful goal for an agent other than maximal reward accumulation. That is, if we detach the notion of value from the agent's environment-specific need at the moment of the decision and instead fix it with respect to some Platonic origin, we can no longer meaningfully speak of satisfying needs, only of optimizing utility.

In this paper, we try to capture the notion of value as the agent's needs by making the valuation process relative to both the agent's current policy and current set of options. This changes the canonical decision modeling approach in three ways: one, it captures the embodied view of cognition; two, we discard the association of numerical rewards to particular outcomes - all rewards are relative comparisons of value; three, decision-making is seen to be a *process* of identifying outcomes from the agent's current situation and evaluating those options relative to the agents understanding of what has constituted the best policy for this situation, based on the agent's past experience.

By formalizing these insights, we show that the assumption of value relativity, while exotic at first sight, can be easily adapted into a tractable choice-learning algorithm. We further show that the choice predictions of this algorithm predict the choices of human subjects across different decision tasks that have heretofore been considered *irrational*. The unforced emergence of a number of previously unconnected cognitive biases from our decision model provides empirical support for its foundational premises. We conclude with a brief discussion of some implications of our findings.

Principles of self-motivated learning

Any realistic model of decision-making must be consonant with the structure of human motivation, i.e., the intrinsic factors that affect an agent's choice behavior. Building upon the relativistic view of the sequential choice selection problem, we now develop a mathematical model of learning choice-selection that appears to be self-motivated and demonstrates realistic learning behavior.

We begin by obtaining alternatives to three assumptions that underpin traditional decision theories - the ontological assumption of an agent-environment duality, the epistemological assumption of absolute numerical utility values and the teleological assumption of utility maximization from our relativistic standpoint.

Embodied representation of preferences

The embodied and embedded view of cognition historically arose as a response to the mainstream computational theory of mind (CTM) that assumes that the mind literally is a digital computer and that thought processes are therefore neural, representational and computational. It suggests that rather than being an isolated observer and computer, the mind acts in inseparable conjunction with the environment to create mental processes (Brooks, 1991). Our relativistic view emphasizes the absence of a unique privileged view of a decision event, and as such, rejects the presence of a disembodied observer. The embodied view of cognition, therefore, lends itself better to our agent-environment description.

Relative utility

Traditional models of cognition have perpetually had an uneasy relationship with quantitative theories of value. The repeated failures of neo-classical economic theories in modern times are traced by behavioral economists to the former's fundamental dependence on a model of reward/value.

Almost all existing decision theories presuppose the existence of state-specific quantified reward. This assumption ignores the failure of the persistent efforts made by early 20th century psychologists towards finding tractable mappings between physical stimuli and the value judgments of human subjects. In fact, it was not until von Neumann and Morgenstern (Neumann & Morgenstern, 1953) showed that it is possible, under a set of mathematical axioms (henceforth the VNM axioms) governing the nature of human preferences, to obtain consistent additive values of relative expected utility among various options that quantifications of human preferences could be meaningfully addressed. The von Neumann program is fundamental to the development of reward-based models of decision-making and planning in AI.

However, two major problems have arisen in the course of the adaptation of the VNM approach to computational decision models. First, it has been established by multiple empirical studies that the VNM axioms do not apply to human preferences. Second, in adapting the VNM approach to computational models, the idea that the additive utilities obtained are relative has been ignored, leading to absolute scalar values of reward unhesitatingly (and errantly!) being used in both the AI and reinforcement learning (Barto & Sutton, 1998) literature.

In our formulation of self-motivated learning, we retreat to the pre-VNM state of understanding of preferences, by assuming that agents can only adopt preferences for particular outcomes relative to others observed in the same context.

Cognitive efficiency

Basic rational choice theory assumes that rational agents attempt to maximize the reward that they can obtain through their actions. However, this assumption has been shown by multiple behavioral studies to be unrealistic. The principal alternative to this assumption is the bounded rationality approach. However, traditional views of bounded rationality (Rubinstein, 2003) continue to assume that agents attempt to maximize reward under computational constraints. From a relativistic standpoint, environmental phenomena are judged to be valuable to the extent they have been judged valuable in the past. Judging utility by *whether* an option has been useful in the past as opposed to *how* useful it is removes the necessity to postulate Platonic rewards embedded in the environment. In our relativistic formulation, the relative value of possible outcomes must emerge from the process of sequential choice selection itself.

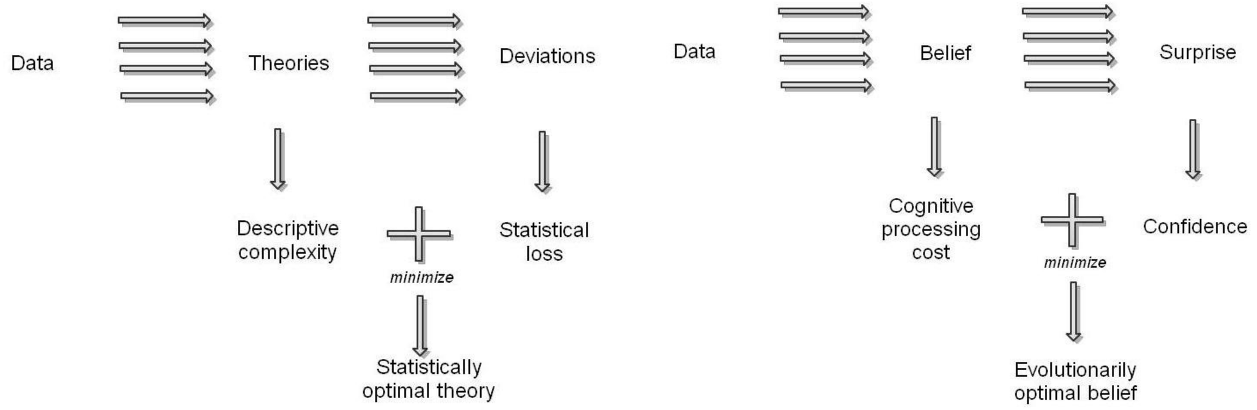
In our formal model, described below, we postulate that humans are essentially searching for minimal-cost theories about how to choose high value options, where the cost is measured in terms of the complexity of encoding and storing the information needed to reliably make these decisions. We term this cost cognitive processing cost, which is equivalent to the cost of accessing past beliefs in the agent's memory.

In brief, rather than view humans and other natural decision-making agents as reward maximizers, or even constrained reward maximizers, we view them as efficient need-satisficers. The expectation of efficiency allows us to pose the optimal control and decision problems using standard optimization techniques, with the relevant quantities to be optimized possessing internal as opposed to external ontological significance.

The decision model

Informally, to make a sequence of decisions, the agent cycles between forming beliefs about the relative worth of options by accessing past experience, termed policies, making choices, experiencing outcomes and updating these policies to minimize processing costs for future decisions.

As we show in Figure 1 and discuss in greater detail in (Srivastava & Schrater, 2010), the formal structure of our model is homologous to the statistical minimum description length principle, with the core premise that an agent tries to minimize its cognitive processing cost T while maintaining a 'satisficingly' high level of predictive confidence C in the quality of its choices. The self-motivated learning objective



(a) A schematic of obtaining a good theory from data following a minimum description length approach.

(b) A schematic of obtaining a good belief from environmental stimuli using an evolutionary MDL principle

Figure 1: This graphic outlines the homologous nature of an evolutionarily optimal meta-cognitive decision strategy with minimum description length principles.

is to minimize a function of the form,

$$\begin{aligned} \underset{\mathbf{x}}{\text{argmin}} \quad & T \\ C_{\text{new}} \quad & \geq C_{\text{old}}. \end{aligned} \tag{1}$$

where T and C are quantified below in terms of policies.

Let the discrete probability distribution $x(s)$ represent an agent's policy, viz., belief about the relative quality of outcomes $s \in \mathcal{S}$ available to it. The surprise experienced by an agent operating with a policy x_a in comparison with policy x_b can be quantified with an information divergence (Kullback & Leibler, 1951) of the form,

$$R(\mathbf{x}_a, \mathbf{x}_b) = \sum_{j=1}^{n_a} \mathbf{x}_a^j(s) \log \frac{\mathbf{x}_a^j(s)}{\mathbf{x}_b^j(s)}. \tag{2}$$

We propose that processing costs are determined by the cost of accessing a belief. Using information-theoretic arguments, we suggest that the access cost of a belief is determined by its predictive exceptionality, which in turn can be measured as a departure from the usual level of surprise that the agent experiences in making its predictions. We measure the informational exceptionality of a past policy x_{old} (and hence the ease with which it will be available for recall to the agent) as the deviation from the average surprise experienced by the agent R':

$$A(\mathbf{x}_{\text{old}}) = |R(\mathbf{x}, \mathbf{x}_{\text{old}}) - \overline{R}|, \tag{3}$$

where x is the agent's current policy.

Given this measure of ease of memory access for each past policy, a reasonable measure of the processing cost of selecting a subset M' out of the set M of all past policies is the inverse availability-weighted sum of the nominal cost of accessing all policies in M'. Assuming the nominal cost of accessing each policy to be unity, the total cost of memory access T becomes,

$$T = \sum_{\mathbf{x}_i \in \mathcal{M}'} A^{-1}(\mathbf{x}_i), \tag{4}$$

Our measure of the agent's confidence in its ability to predict its environment, $C : x \to [0, 1]$ captures the idea that confidence grows when the policies have low uncertainty and low surprise:

$$C = \frac{1}{C_{\max}} \frac{\log |\mathbf{x}| - H(\mathbf{x})}{\sum_{\text{memory}} R(\mathbf{x}, \mathbf{x}_{\text{old}})}, \tag{5}$$

where the numerator is a monotonically decreasing function of the Shannon entropy $H(x)$ of the policy. Note that C is normalized with respect to the greatest value it has previously been observed to achieve.

Any algorithmic solution of our agent's objective function must solve three problems - one, specify a memory update specifying how existing policies are integrated into the agent's current policy; two, specify an environmental update, which shows how the perceived goodness of various outcomes at the present moment, which we call *reward-inference*[1], are integrated into the agent's current policy; three, specify a combinatorial optimization algorithm specifying which subset of existing policies the agent will *recall* to form its new policy, such that the objective function we have defined above is satisfied. In (Srivastava & Schrater, 2010), we present a direct policy search-based solution to all three problems for simple outcome spaces, resulting in a self-motivated learning algorithm for predicting choices made by agents in sequential settings. The resultant algorithm outputs

[1] Our model assumes that sensory data is encoded into the space of possible outcomes as a relative preference by existing neuronal processes. Thus, our usage of the term *reward-inference* accentuates the fact that it is obtained after perceptual processing of environmental stimuli.

beliefs corresponding to the relative preference for each of the possible outcomes in the agent's decision context.

While simply constructing a choice-learning algorithm capable of quantifying internal motivations would pass for an interesting theoretical exercise, in the next section, we present evidence below to show that our approach makes strong falsifiable predictions about classes of behaviors in human subjects that classical approaches are unable to explain.

Irrational behavior is rational

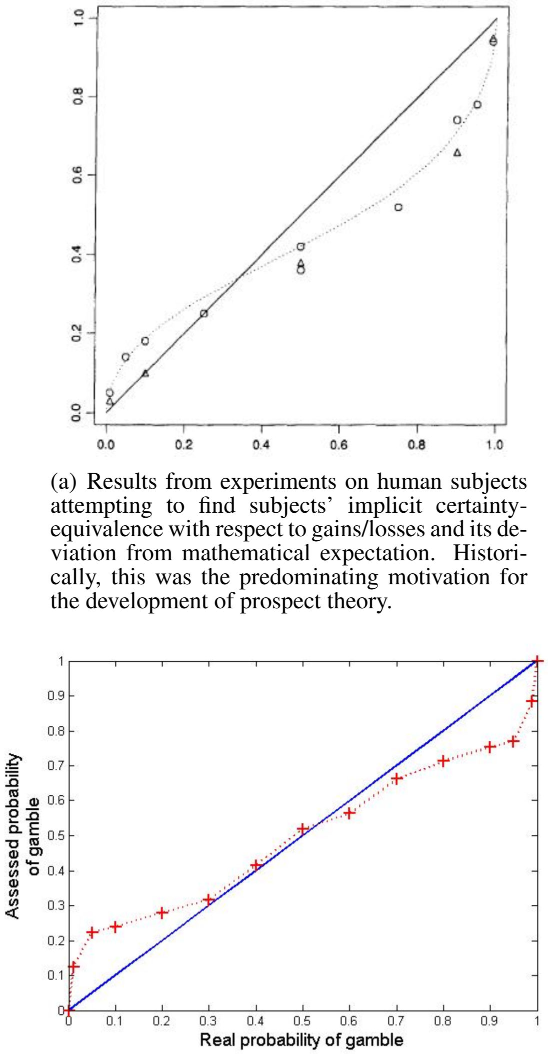
Counter-examples to the expected utility axioms have a history nearly as old as rational choice theory itself. Deviations from the predictions of expected utility theories, when consistently observed in human subjects under controlled experimental settings, have been defined as cognitive 'biases', implying a somewhat paternalistic premise that people would behave the way expected utility theory predicts if they were unbiased (smart) enough. However, the epistemic value of decision theories that are consistently mistaken cannot be reasonably justified[2].

While the weakness of existing neo-classical decision models have been critiqued exhaustively in the behavioral economics literature, incremental fixes to the basic framework and the use of ad hoc heuristics has created a fragmented universe of predictive models, each successful at explaining one specific set of cognitive biases. We believe that a causally valid theory of decision-making must necessarily present a unified explanation for multiple families of cognitive biases.

To demonstrate the plausibility of our self-motivated decision model, we show how its predictions replicate biases observed in two different cognitive domains and heretofore explained separately using generalized utility theories in one case and evolutionary heuristics in the other. Similar results for other families of biases are described in a longer technical report (Srivastava & Schrater, 2010), but are omitted in this paper due to space constraints. We suggest that the unforced emergence of multiple classes of cognitive biases from our model provides support for its cognitive realism.

Risk aversion

Kahneman and Tversky (Tversky & Kahneman, 1992) proposed prospect theory largely to explain deviations from expected value predictions in certainty-equivalence studies on evaluations of risky prospects in human subjects. They observed that subjects consistently exhibited a four-fold pattern of behavior when confronted with risk: risk seeking for gains with low probability, risk aversion for gains with high probability, risk seeking for losses with high probability, and risk aversion for losses with low probability. (Tversky & Kahneman, 1973) explain the emergence of this pattern as a consequence of the disproportionate weighting of low-probability outcomes in human subjects. This explanation was subsequently amended in (Tversky & Kahneman, 1992) to restrict over-weighting only to 'extreme' low-probability events as opposed to all low probability events.



(a) Results from experiments on human subjects attempting to find subjects' implicit certainty-equivalence with respect to gains/losses and its deviation from mathematical expectation. Historically, this was the predominating motivation for the development of prospect theory.

(b) Results from simulation of prospect theory experiment using existential agents as subjects. The blue line represents the idealized expected value prediction while the red markers indicate average preference of 200 agents having experienced a history of repeated exposure to a choice selection task between a risky gamble with a certain (x-axis) probability of succeeding and a safe choice.

Figure 2: Existential learning generatively reproduces experimental results previously explained only empirically using prospect theory.

The experimental setup for their experiments is fairly straightforward: subjects are asked to select between a 'safe' gain/loss prospect of known value and one of unknown value determined as a Bernoulli choice between two known outcomes. For example, a subject could be asked to choose between selecting a prospect that pays $0 with a probability of

[2] The dismissal of various cognitive phenomena as 'biases' is rather reminiscent of the construction of epicycles in Ptolemaic astronomy - it is not the theory that does not fit the observations, it is the observations that are so inconsiderate that they don't obediently fit the theory!

0.9 and $50 with a probability of 0.1 and a set of prospects guaranteed to pay anywhere between $2 and $20 (say). The subjects were required to indicate their preference between the risky and safe prospects for all the safe prospects presented to them. The certainty equivalent value was estimated as the midpoint between the lowest accepted and the highest rejected value from among the safe prospects. Selections where the certainty equivalent value exceeded the expected value of the risky prospect ($5 in this case) were considered risk-seeking, while those that were lower were counted as risk averse.

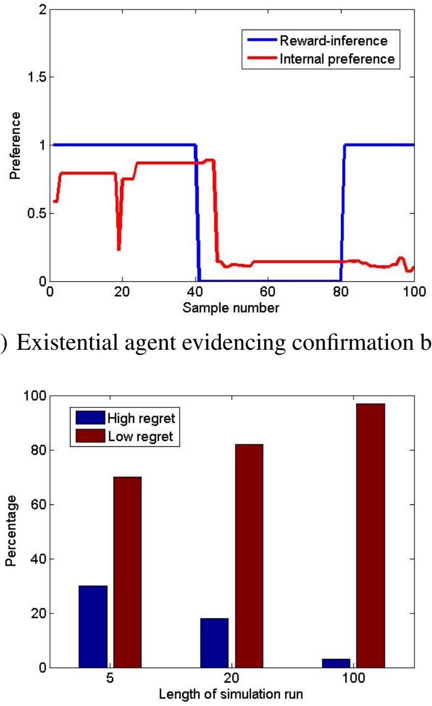
In order to simulate the experimental setup described in (Tversky & Kahneman, 1992), we design our outcome space to consist of two possible outcomes: select safe prospect or select risky prospect. For every decision instance, the payoff for the risky prospect is sampled from a Bernoulli distribution appropriate for the gamble. For the gamble in the example above, this means that the risky prospect will pay $0 in about 9 out of every 10 decision instances. The reward-inference signal is constructed to assign a preference of 1 to the better prospect (and 0 to the worse prospect) at every instantiation. Thus, a choice between a gamble with a 0.1 probability of paying off against a certain safe outcome is modeled as a generative mechanism for reward-inference that reflects a selection $\{0\ 1\}$ biased towards the safe choice 90% of the time and the alternate risky choice $\{1\ 0\}$ 10% percent of the time.

We provided each one of a population of 200 agents with a series of 100 such reward-inference signals. A series is presumed to indicate the ‘learning’ phase for an agent with respect to a particular choice problem involving risk evaluation. At the end of a series, the agent is assumed to possess, in the form of its final preference, an evaluative model for selecting between the prospects offered in the (Tversky & Kahneman, 1992) selection task. We modify the probability of winning or losing the gamble by modifying the Bernoulli distribution parameterizing the reward inference distribution.

In Figure 2, we see that our simulation replicates results that are qualitatively similar to the experimental results obtained from human subjects in (Tversky & Kahneman, 1992). Remarkably, agents running our existential learning algorithm consistently present the same four-fold pattern of risk aversion observed in human subjects. This leads us to hypothesize that the biases documented by Kahneman and Tversky, which have subsequently motivated the development of prospect theory and other generalized expected utility theories are, in fact, adaptive in nature rather than existing a priori in human decision-makers. Our model produces, to the best of our knowledge, the first generative mechanism for estimating and potentially quantifying Kahneman and Tversky’s four-fold pattern of risk aversion.

Confirmation biases

The term ‘confirmation bias’ often references biased hypothesis evaluation, differential memory recall, belief divergence, attitude polarization and other biases arising in different experimental contexts. The fundamental similarity shared by all these biases is the tendency for subjects to prefer information that confirms their existing preconceptions/hypotheses over objective evidence. Confirmation biases are generally explained using availability or priority heuristics in the biases and heuristics literature (Ayal & Hochman, 2009). Rational choice theories find it difficult to account for their presence.



(a) Existential agent evidencing confirmation bias

(b) Percentage of high and low surprise instances active in agent’s salient set for runs of different lengths.

Figure 3: Different flavors of confirmation bias exhibited by existential learning algorithm.

Figure 3(a) displays typical performance of an existential agent on a binary prediction task. Given consistent reward-inference favoring one outcome (say 0, 1), the agent’s preference for this outcome increases, which is entirely rational. Then, consistent reward-inference favoring the other outcome 1, 0 is provided, causing the agent to reverse its preference (after a brief delay), which, again is entirely rational.

However, when the reward-inference is switched yet again back to the original outcome, the agent does not switch, but continues to confirm its recent preference for the other outcome. This superficially irrational behavior follows naturally from the tendency of our agent to retain its existing theory if formulating a newer theory would cause its predictive confidence to drop.

The first scientific evaluation of confirmation bias is historically assigned to Wason’s (Wason, 1960) rule-discovery experiments. However, Klayman and Ha (Klayman & Ha, 1987) proved that what Wason had actually shown was that human subjects prefer using positive test strategies, i.e., instead of trying to find counter-examples to a hypothesis, they seek to validate it. Interpreting these findings in our framework, observe that a disconfirming negative test strategy of trying to rigorously disprove a held hypothesis would create several high surprise/regret decision instances for an existential agent.

Conversely, deploying positive test strategies would create (given a predictable environment) low surprise/regret instances. Since part of the agent's existential goal is to maximize its expectation of future reward, and since this expectation, in the form of confidence, will vary inversely with the cumulative surprise in the agent's recalled history, it will strongly prefer making choices that lead to low surprise, and hence will prospectively prefer positive test strategies. Figure 3(b) shows the existential agent's preference for low surprise decision instances.

Very interestingly, we find that the agent's preference for positive test strategies appears to emerge gradually as it becomes surer of its existing hypothesis. This corroborates the information-theoretic intuition (Klayman & Ha, 1987) that such a preference arises as an information-processing response to environments where positive queries have higher informational content than negative queries.

Discussion

In this work, we have showed how two different families of cognitive biases can, in fact, be generated from a single causal model of decision-making, merely by shifting the objective of a classical bounded rational agent from resource-constrained utility maximization to prediction-constrained cognitive effort minimization. This change in perspective, in turn, is obtained from a relativized view of the nature of preferences, arising from the intuition that, however an agent may be engaged with its environment, it sees, from its own existentially stationary vantage point, a set of options that it has seen in the past, recalls its previous experience of choice selection amongst them, and uses its memory recall to make a new choice selection. Thus, the principal quantity of interest in decision-making becomes the cost of memory recall, optimizing which results in a novel rational decision theory.

Note further that, by retaining a sense of optimality, we have essentially proposed a new way of defining rational utility, which subsumes positive aspects of both the classical expected utility paradigm (Neumann & Morgenstern, 1953) and recent heuristic-based methods (Ayal & Hochman, 2009) while avoiding their defects. Specifically, our model retains the analytical tractability and causal interpretability of the traditional expected utility/rational choice paradigm while adapting the definition of rationality to confirm with Gigerenzer's (Gigerenzer & Goldstein, 1996) idea of 'ecological rationality'. By adopting an embodied representation of the agent-environment interface, and an information-theoretic basis for defining costs, we are able to generalize our model's dependence across the ecology of different domains. This allows our model to retain predictive accuracy both in typical environmental settings, where it replicates the predictions of classical rational choice models, as well as in atypical environmental settings, where it faithfully replicates biases observed in human subjects. As a consequence, our results show that a number of important cognitive biases emerge as natural consequences of the way a metacognitive agent encodes information about the environment. However, our model is as yet unable to account for set size and framing effects, since it currently considers a particular embodied representation of the agent-environment as persistent across choices. Extending our model to capture these effects is an exciting direction for future work, with cross-disciplinary implications.

Références

Akerlof, G., & Shiller, R. (2009). *Animal spirits: How human psychology drives the economy, and why it matters for global capitalism.* Princeton University Press.

Ariely, D. (2009). *Predictably irrational: The hidden forces that shape our decisions.* Harper Collins.

Ayal, S., & Hochman, G. (2009). Ignorance or integration: The cognitive processes underlying choice behavior. *Journal of Behavioral Decision Making*, *22*, 455-474.

Barto, A., & Sutton, R. (1998). *Reinforcement learning: an introduction.* Univesity of Cambridge Press.

Brooks, R. (1991). Intelligence without representation. *Artificial Intelligence*, *47*, 139-159.

Gigerenzer, G., & Goldstein, D. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychological Review*, *103*(4), 650-669.

Klayman, J., & Ha, Y.-W. (1987). Confirmation, disconfirmation and information in hypothesis testing. *Psychological Review*, *94*, 211-228.

Kullback, S., & Leibler, R. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, *22*, 79-86.

Mill, J. (1874). *Essays on some unsettled questions of political economy, par 5.38.*

Neumann, J., & Morgenstern, O. (1953). *Theory of games and economic behavior.* Princeton University Press.

Persky, J. (1995). Retrospectives: The ethology of homo economicus. *The Journal of Economic Perspectives*, *9*(2), 221-231.

Rubinstein, A. (2003). *Modeling bounded rationality.* Prentice-Hall.

Srivastava, N., & Schrater, P. (2010). *An evolutionarily motivated model of decision-making under uncertainty.* Available at SSRN: http://ssrn.com/abstract=1687205.

Tversky, A., & Kahneman, D. (1973). Availability: a heuristic for judging frequency and probability. *Cognitive Psychology*, *5*, 207-232.

Tversky, A., & Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*, 297-323.

Wason, P. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, *12*, 129-140.