Finding Foundations for Bounded & Adaptive Rationality

Workshop Program

Max Planck Institute for Human Development

Friday 22 - Sunday 24
November 2013

Workshop Organizers: Ralph Herwig
Arthur Paul Pedersen
Renata Suter

Center for Adaptive Rationality | Max Planck Institute for Human Development

Travel, Hotel, Board, and Funding Coordinator Manuela Meermann
Restaurant Coordinator and Funding Assistance Yvonne Bennett
Poster Designer Mauren Amelia Antkowski

Funding for the workshop has been provided by the German Research Foundation and the Max Planck Institute for Human Development.

Imprint

Copyeditor Arthur Paul Pedersen
Program Layout and Design Dagmar Gülow
Cover Design Mauren Amelia Antkowski

Max Planck Institute for Human Development
Lentzealle 94, 14195 Berlin
www.mpib-berlin.mpg.de
info@mpib-berlin.mpg.de

Max-Planck-Institut für Bildungsforschung
Max Planck Institute for Human Development

NEW • FRAMEWORKS • OF
RATIONALITY
DFG • SCHWERPUNKTPROGRAMM • 1516

Max-Planck-Institut für Bildungsforschung
Max Planck Institute for Human Development

Center for Adaptive Rationality

Finding Foundations for Bounded and Adaptive Rationality

22-24 November 2013 Workshop Program

Funding for the workshop has been provided by the German Research Foundation and the Max Planck Institute for Human Development.

Organizers

Ralph Hertwig
Arthur Paul Pedersen
Renata Suter

Friday 22 November

9:15 AM	*Workshop Agenda:* **Ralph Hertwig & Arthur Paul Pedersen**
9:55 AM	*Welcome Reception*
10:15 AM	**Peter Todd** Building the Theory of Ecological Rationality (*with Henry Brighton*)
10:45 AM	Discussant: Jonathan Nelson
11:00 AM	Discussant: Michael Bishop
11:15 AM	*Public Discussion*
11:45 PM	*Lunch* (in the Cantine)
1:30 PM	**Jason Dana** Advances in Improper Linear Models Research and How They Inform Decision Making Theories (*with Clintin Davis-Stober*)
2:00 PM	Discussant: Wai-Tat Fu
2:15 PM	Discussant: Mirjam Jenny
2:30 PM	*Public Discussion*
3:00 PM	*Coffee Break*
3:30 PM	**Patrick Suppes** Qualitative Axioms of Uncertainty as a Foundation for Probability and Decision Making
4:00 PM	Discussant: Gerhard Schurz
4:30 PM	*Public Discussion*
5:00 PM	*Break*
6:00 PM	*Dinner* (at the Max Planck Institute for Human Development)

Max-Planck-Institut für Bildungsforschung
Max Planck Institute for Human Development

Center for Adaptive Rationality

Finding Foundations for Bounded and Adaptive Rationality

22-24 November 2013 Workshop Program

Saturday	23 November
9:00 AM	**Thomas Sturm** Naturalistic Epistemology and Adaptive Accounts of Rationality
9:30 AM	Discussant: Sarah Wellen
9:45 AM	Discussant: Michael Waldmann
10:00 AM	*Public Discussion*
10:30 AM	*Coffee Break*
11:00 AM	**Wai-Tat Fu** The Central Role of Cognitive Search in Bounded and Adaptive Rationality
11:30 AM	Discussant: Özgür Simsek
11:45 AM	Discussant: David Kellen
12:00 PM	*Public Discussion*
12:30 PM	*Lunch* (at the Max Planck Institute for Human Development)
2:00 PM	**Gerhard Schurz** Strategy-Selection by Meta-Induction in Prediction and Decision Tasks under Changing Environments: Theoretical and Empirical Results (*with Paul Thorn and Christian Feldbacher*)
2:30 PM	Discussant: Thomas Sturm
2:45 PM	Discussant: Konstantinos Katsikopoulos
3:00 PM	*Public Discussion*
3:30 PM	*Coffee Break*
4:00 PM	**Hansjörg Neth** Rational Task Analysis (*with Chris R. Sims and Wayne D. Gray*)
4:30 PM	Discussant: Nadine Fleischhut
4:45 PM	Discussant: Till Grüne-Yanoff
5:00 PM	*Public Discussion*
5:30 PM	*Break*
6:00 PM	*Dinner* (at Piazza-Michelangelo)

Center for Adaptive Rationality

Finding Foundations for Bounded and Adaptive Rationality

22-24 November 2013 Workshop Program

Sunday 24 November

10:00 AM	**Sarah Wellen** Fast and Frugal Learning (*with David Danks*)
10:30 AM	Discussant: Stefan Herzog
10:45 AM	Discussant: Ralf Mayrhofer
11:00 AM	*Public Discussion*
11:30 AM	*Coffee Break*
12:00 PM	**Michael Bishop** The Benefits of an Inclusive Approach to the Study of Rationality (*with Heather Cipolletti*)
12:30 PM	Discussant: David Danks
12:45 PM	Discussant: Björn Meder
1:00 PM	*Public Discussion*
1:30 PM	*Lunch* (at the Max Planck Institute for Human Development)
3:00 PM	*Final Comments*: **Ralph Hertwig & Arthur Paul Pedersen**
3:30 PM	*Final Discussion*
5:00 PM	*Farewell*

Speaker and Commentators

Michael Bishop	Philosophy, Florida State University,
Jason Dana	Psychology, University of Pennsylvania
David Danks	Philosophy, Carnegie Mellon University
Nadine Fleischhut	Center for Adaptive Rationality, MPI for Human Development
Wai-Tat Fu	Computer Science, University of Illinois at Urbana-Champaign
Till Grüne-Yanoff	Adaptive Behavior and Cognition, MPI for Human Development
Stefan Herzog	Center for Adaptive Rationality, MPI for Human Development
Mirjam Jenny	Center for Adaptive Rationality, MPI for Human Development
Konstantinos Katsikopoulos	Adaptive Behavior and Cognition, MPI for Human Development
David Kellen	Center for Adaptive Rationality, MPI for Human Development
Ralf Mayrhofer	Psychology, University of Göttingen
Björn Meder	Adaptive Behavior and Cognition, MPI for Human Development
Jonathan Nelson	Adaptive Behavior and Cognition, MPI for Human Development
Hansjörg Neth	Adaptive Behavior and Cognition, MPI for Human Development
Gerhard Schurz	Philosophy, Heinrich Heine University Düsseldorf
Özgür Simsek	Adaptive Behavior and Cognition, MPI for Human Development
Thomas Sturm	Philosophy, Autonomous University of Barcelona
Patrick Suppes	Philosophy, Stanford University
Peter Todd	Psychological and Brain Sciences, Indiana University Bloomington
Michael Waldmann	Psychology, University of Göttingen
Sarah Wellen	Philosophy, Carnegie Mellon University

Building the Theory of Ecological Rationality

Peter M. Todd
Henry Brighton

Friday 22 November, 10:15 AM

Research perspectives on decision making have often fallen at one or the other end of Archilochus's famous dichotomy between hedgehogs and foxes: having one big idea to approach all the decisions in life, or many separate ideas for dealing with different kinds of decisions. The former hedgehogian approach corresponds to having one powerful theory (e.g., subjective expected utility, Bayesian probability updating) to predict and explain the full range of decision making, while the latter foxian approach manifests as a typically atheoretical collection of different decision-making strategies for different situations (e.g., a bag of biased heuristics). The research program of ecological rationality aims to bridge these two perspectives by providing a theoretically-driven account of the variety of different decision mechanisms people draw on. This hedgefoxian account (after Loewenstein, Vohs, & Baumeister, 2007) explains the effectiveness of particular decision heuristics in terms of their fit to the structures of information available in the task environment. But it risks also devolving into a collection of unconnected mind-environment match-ups if the theoretical foundations predicting when different mental structures will fit with different environmental structures are not well-developed. Here we discuss the current state of theorizing regarding ecological rationality, and describe ways forward. We highlight both a top-down approach in which powerful machine learning models are constrained and simplified through the introduction of specific biases that make the models more robust in particular types of environments via the bias-variance tradeoff, and a bottom-up approach that starts with psychologically plausible building blocks that can be combined to create simple heuristics that are also fit for specific environments.

Reference

Loewenstein, G., Vohs, K. D., & Baumeister, R. F. (2007). Introduction: The hedgefox. In K. D. Vohs, R. F. Baumeister, & G. Loewenstein (Eds.), *Do Emotions Help Or Hurt Decision Making? A Hedgefoxian Perspective* (pp. 3–9). New York: Russell Sage Foundation.

Advances in Improper Linear Models Research and How They Inform Decision-Making Theories

Jason Dana
Clintin Davis-Stober

Friday 22 November, 1:30 PM

Fixed, *a priori* weighting schemes for the linear model, such as equally weighting all predictors, can be surprisingly accurate on cross-validation when compared with regression approaches. We review recent advances in the understanding of when and why such "improper" linear models work. Specifically, this work recasts fixed weighting schemes as estimates of the population coefficients and maps the circumstances under which they are good estimates. We extend this work to the understanding of fast and frugal heuristics. We argue that many such heuristics can be represented within the linear model framework, allowing us to map the environments in which particular heuristics approach optimal performance. In this way, we can predict what heuristics an adaptive decision maker will use given the features of a decision environment. Conversely, given that a decision maker uses a given

heuristic in some environment, we can formulate hypotheses about the decision maker's objective function. We find that familiar heuristics have a rational underpinning and, in fact, closely mimic statistical optimization.

Qualitative Axioms of Uncertainty as a Foundation for Probability and Decision Making

Patrick Suppes **Friday 22 November, 3:30 PM**

Although the concept of uncertainty is as old as Epicurus's writings, and an excellent quantitative theory, with entropy as the measure of uncertainty having been developed in recent times, there has been little exploration of the qualitative theory.

The purpose of the present paper is to give a qualitative axiomatization of uncertainty, in the spirit of the many studies of qualitative comparative probability. The qualitative axioms are fundamentally about the uncertainty of a partition of the probability space of events. Of course, it is common to speak of the uncertainty, or randomness, of a random variable, but only the partition defined by the values of the random variable enter into the definition of uncertainty, not the actual values.

It is straightforward to add axioms for decision making following the general line of Savage from the 1950s. Indeed, in the spirit of Epicurus, it is really our intuitive feeling about the uncertainty of the future that motivates much of our thinking about decisions. Here, the distinction between the concepts of probability and uncertainty can be made by citing many familiar examples.

Without spelling out the technical details, the axiomatization of qualitative probability with uncertainty as the most important primitive concept, it is possible to raise a different kind of question about bounded rationality. This new question is whether or not one should bound the uncertainty in thinking and investigating any detailed framework of decision making. Discussion of this point is certainly different from the question of bounding rationality by not maximizing expected utility. In practice, we naturally bound uncertainty in our analysis of decision-making problems. As in the case of formulating an alternative for maximizing expected utility, so is the case of rational alternatives to maximizing uncertainty. There are several issues to consider. In the spirit of my other work in qualitative probability, I explore alternatives rather than attempt to give a definitive argument for one single solution.

Naturalistic Epistemology and Adaptive Accounts of Rationality

Thomas Sturm **Saturday 23 November, 9:00 AM**

Naturalistic epistemology, broadly conceived, is the view that the philosophical theory of knowledge should be based on results and methods of the empirical sciences. What do adaptive accounts of human reasoning have to offer here? Stated differently, what happens if the idea that human rationality is somehow a product of evolution and in some sense adaptive meets the quest for an improvement of our epistemic undertakings? I shall discuss two relevant approaches.

On the one hand, there is the strong version of an evolutionary psychology of rationality (EPR)—strong because it not only asserts that rationality has evolved somehow but, more specifically, that evolution results in the emergence of strictly domain-specific reasoning modules that are the result of *adaptations* (e.g., Cosmides, 1989; Cosmides & Tooby, 1996; Barkow, Cosmides & Tooby, 1992; Sugiyama, Tooby & Cosmides, 2002; cf. also Confer et al. 2010). On the other hand, there is the program of bounded rationality (BR), according to which human reasoning is content-and-context dependent, and this in a both descriptive and in a normative sense as well (see, e.g., Gigerenzer, 2000; Gigerenzer & Goldstein, 1996; Gigerenzer & Selten, 2001; Gigerenzer, Todd & the ABC Group, 1999).

Certainly there are affinities between the two approaches. For instance: (1) Both start from discussions over how to deal with cases of allegedly bad reasoning—for example, in the Wason selection task, the Linda problem, in tests concerning the ability to use proper sample sizes, and so on. (2) Both combat common enemies, namely the heuristics-and-biases approach of Kahneman, Tversky, and others, as well as optimizing and overly formalistic notions of rationality (e.g., Kahneman, Slovic & Tversky, 1982). Finally, (3) both use the vocabulary of adaptation and related Darwinian notions (for the bounded rationalists, see, e.g., Gigerenzer, 2001; Gigerenzer, Todd & the ABC Group, 1999, e.g., p. 30; Todd, Hertwig & Hoffrage, 2005). They have therefore even been viewed as pursuing the same research program (e.g., Clark, 2001; Samuels, Stich & Bishop, 2002; Stanovich & West, 2003; Stenning & Van Lambalgen, 2008). Contrary to this, I shall argue that the two approaches differ over fundamental assumptions, and that BR fits better with the goal of naturalizing epistemology than EPR does. It seems, however, that there is a price to be paid for this.

In Section I, I identify relevant differences between the two programs. These concern two central and related theoretical views concerning the entangled issues of modularity and adaptivity. (i) EPR commits itself to the "massive modularity thesis." In contrast, defenders of BR need not and should not do this—despite the fact that they favor content-and-context-dependent rules of reasoning. In fact, they frequently assert that fast and frugal heuristics are applied across (and work well in) quite different domains or reasoning tasks. (ii) Bounded rationalists need not commit themselves to a deeply evolutionary notion of adaptation. Their notion of rationality is an *ecological* one. Although the application of a heuristic is not a one-shot game, and its performance in repeated situations should probably affect its actual occurrence (Arnau, Ayala & Sturm, 2014), BR is not committed to disputed claims about the selection and inheritance of reasoning strategies, or what may have happened in the environment of evolutionary adaptation. What primarily matters for BR is the relation between heuristics and reasoning problems (and/or environments).

In section II, I introduce relevant versions of naturalized epistemology. Some, such as Quine's early (1969) "replacement naturalism" or the evolutionary epistemology of the 1970s (e.g., Vollmer, 1975), tried to *replace* standard epistemological questions by descriptive or explanatory ones. But this is a dead end for naturalized epistemology (Kim, 1988). Naturalistic epistemology needs to be normative. Now, since the theory of knowledge and the theory of rationality have genuine tasks of their own, we cannot without further ado move from a normative naturalism about rationality (e.g., Gigerenzer & Sturm, 2012) forward to a normative naturalism about epistemology. Yet, there is an important area where epistemology and theories of rationality overlap: the improvement of our beliefs. Psychological theories that give adequate advice for how to revise or improve our beliefs through better reasoning are plausible candidates. Yet, even naturalistic epistemologists who have taken this normative road have not really considered what role adaptive accounts of rationality might play in this context. More typically, they have primarily considered the heuristics and biases approach (e.g., Goldman, 1986, 2008; Kitcher, 1992; Stich, 1980, 1990; for partial exceptions, see Kornblith, 1993; Bishop & Trout, 2005). Moreover, they have typically left their favorite psychological approach uncriticized.

In section III, I first address criticisms of EPR and then suggest an argument for why BR is more consistent with a normative epistemology. I agree with several critics of EPR: (a) The evidence for domain-specific reasoning modules is unconvincing—for example, concerning the Wason selection task, Cosmides and Tooby fail to take due care of the purely *formal* difference between descriptive and deontic conditionals (Fodor, 2000, 2008; Stenning & Van Lambalgen, 2008; attempts to rebut this point by Beaman, 2002, Cosmides, Tooby, Fiddick & Bryant, 2005, or Cosmides & Tooby, 2008 seem unsuccessful to me). (b) Cosmides and Tooby have also failed to furnish cogent grounds for their claims about the adaptivity of the alleged reasoning modules (Richardson, 2005). Proponents of BR as well have noted in a similar vein that one cannot infer from the fact that a heuristic works well in a certain environment that it must be an adaptation to that environment (Hutchinson & Gigerenzer, 2005). But the objection I wish to use is a different one: (c) While EPR might identify some innate constraints on reasoning, it is in strong tension with normative epistemology. If the claims of massive modularity and adaptation (due to natural selection) were true, they would undermine a normative naturalistic epistemology. Quine claimed that, in principle, all our beliefs are open to revision. One need not go as far as that, but improving our web of beliefs presupposes that, to a sufficient degree at least, our reasoning processes are not too isolated from one another, or operate in an integrated fashion (Over, 2002). More importantly, they must also be open to *responsible* evaluation and revision. The BR approach is more consistent with such ideas: It does not assume the massive modularity thesis nor does it claim that heuristics are always domain-specific. Moreover, defenders of BR are aware of the necessity of critical discussions, and the limits, of their own favorite heuristics. Since environments and/or reasoning problems and constraints may (nowadays at least) change quickly, reasoners should in principle be able to consciously and reflectively *choose* from a variety of strategies. (That they should often but not always favor satisficing over optimization is one instance of such a recommendation.) The price to be paid is that talk of adaptivity in the BR program is so markedly different from the one used in evolutionary psychology or biology that it would be better to abandon it. In work on BR, "adaptive" simply means that reasoning strategies work in a certain environment, or for certain reasoning tasks (or a certain range thereof). Standard epistemic notions, such as reliability or truth-conduciveness, will do the job just as well, and they will free BR from frequent misunderstandings. However, the issue is not merely terminological: The link between biological evolution and BR will have to be reconsidered too. For instance, how can it be explained that heuristics did evolve such that they can be applied to different reasoning tasks or domains?

Of course, objection (c) against EPR might look like a case of the moralistic fallacy ("X ought not to be the case, so X cannot be the case"). But this isn't so. After all, it is a serious issue, both conceptually and empirically, whether reasoning strategies are content-and-context dependent, organized in a modular fashion, and so on. And clearly, these concepts need to be distinguished more clearly from the mind-environment dependency thesis of BR. Finally, it is conceivable that even highly general reasoning strategies possess an evolutionary function (although proposals in that direction such as Nozick's (1993) are criticized for being vague (Richardson, 2007)). We may therefore use the tension between EPR and normative epistemology as a tool for clarifying and improving fundamental conceptual assumptions of BR. And since epistemology should not blindly accept BR's basic approach and its results, we might speak of a "critical normative naturalized epistemology"—if only that wasn't such an ugly expression.

References

Arnau, E., Ayala, S., & Sturm, T. (2014). Cognitive externalism meets bounded rationality. *Philosophical Psychology, 27,* 50–64.

Barkow, J. H., Cosmides, L., & Tooby, J. (Eds.). (1992). *The adapted mind: Evolutionary psychology and the generation of culture.* Oxford: Oxford University Press.

Beaman. (2002). Why are so good at catching cheaters. *Cognition, 83,* 215–220.

Bishop, M. A., & Trout, J. D. (2005). *Epistemology and the psychology of human judgment.* New York: Oxford University Press.

Buller, D. (2005). Evolutionary psychology: The emperor's new paradigm. *Trends in Cognitive Sciences, 9*(6), 277–283.

Buller, D. (2006). *Adapting minds: Evolutionary psychology and the persistent quest for human nature.* Cambridge, MA: MIT Press.

Buller, D. (2008). Four fallacies of pop evolutionary psychology. *Scientific American, 300*(1), 74–81.

Buller, D., Fodor, J., & Crume, T. L. (2005). The emperor is still underdressed. *Trends in Cognitive Sciences, 9,* 508-510.

Clark, A. (2001). Reasons, robots, and the extended mind. *Mind & Language, 16,* 121–145.

Confer, J., Easton, J. A., Fleischman, D., Goetz, C., Lewis, D., Perrilloux, C., & Buss, D. (2010). Evolutionary psychology: Controversies, questions, prospects, and limitations. *American Psychologist, 65,* 110–125.

Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition, 31,*187–276.

Cosmides, L., & Tooby, J. (1996). Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty. *Cognition, 58,* 1-73.

Cosmides, L., Tooby, J., Fiddick, L., & Bryant, G. A. (2005). Detecting cheaters. *Trends in Cognitive Sciences, 9,* 505–506.

Cosmides, L., & Tooby, J. (2008). When falsification strikes: A reply to Fodor. In W. Sinnott-Armstrong (Ed.), *Moral psychology, Volume 1: The evolution of morality: Adaptations and innateness* (pp. 143–161). Cambridge, MA: Bradford Books.

Fodor, J. (2000). Why we are so good at catching cheaters. *Cognition, 75,* 29–32.

Fodor, J. (2008). Comment on Cosmides and Tooby. In W. Sinnott-Armstrong (Ed.), *Moral psychology, Volume 1: The evolution of morality: Adaptations and innateness* (pp. 137–142). Cambridge, MA: Bradford Books.

Gigerenzer, G. (2000). *Adaptive thinking.* New York: Oxford University Press.

Gigerenzer, G. (2001). The adaptive toolbox: Towards a Darwinian rationality. In J. A. French, A. C. Kamil, & D. W. Leger (Eds.), *Evolutionary psychology and motivation* (Nebraska Symposium on Motivation, Vol. 48, pp. 113–143). Lincoln: University of Nebraska Press.

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

Gigerenzer, G., & Hug, K. (1992). Domain-specific reasoning: Social contracts, cheating, and perspective change. *Cognition, 43,* 127–171.

Gigerenzer, G. & Selten, R. 2001. Bounded rationality: The adaptive toolbox. London: MIT Press.

Gigerenzer, G., Todd, P. M., & the ABC Research Group. (1999). *Simple heuristics that make us smart.* New York: Oxford University Press.

Gigerenzer, G., & Sturm, T. (2012). How (far) can rationality be naturalized? *Synthese.*

Goldman, A. (1986). *Epistemology and cognition.* Cambridge, MA: Harvard University Press.

Goldman, A. (2008). Human rationality: Epistemological and psychological perspectives. In: A. Beckermann & Sven Walter (Eds.), *Philosophie: Grundlagen und Anwendungen* (pp. 230-247). Paderborn: Mentis.

Hutchinson, J., & Gigerenzer, G. (2005). Simple heuristics and rules of thumb: Where psychologists and behavioural biologists might meet. *Behavioural Processes, 69,* 97–124.

Kahneman, D., Slovic, P., & Tversky, A. (1982). (Eds.). *Judgment under uncertainty: Heuristics and biases.* New York: Cambridge University Press.

Kim, J. (1988). What is "naturalized epistemology?" *Philosophical Perspectives, 2,* 381–405.

Kitcher, P. (1992). The naturalists return. *Philosophical Review, 101,* 53–114.

Kornblith, H. (1993). Inductive inference and its natural ground. Cambridge, MA: MIT Press.

Nozick, R. (1993). *The nature of rationality.* Princeton: Princeton University Press.

Over, D. (2002). The rationality of evolutionary psychology. In J. L. Bermúdez & A. Millar (Eds.), *Reason and nature: Essays on the theory of rationality* (pp. 187–207). Oxford: Clarendon.

Quine, W. V. O. (1969). Epistemology naturalized. In W. V. O. Quine, *Ontological relativity and other essays* (pp. 69–90). New York: Columbia Press.

Richardson, R. (2007). *Evolutionary psychology as maladapted psychology.* Cambridge, MA & London: MIT Press.

Samuels, R., Stich, S., & Bishop, M. (2002). Ending the rationality wars: How to make disputes about human rationality disappear. In R. Elio (Ed.), *Common sense, reasoning and rationality* (pp. 236–268). New York: Oxford University Press.
Stanovich, K. E., & West, R. F. (2003). Evolutionary versus instrumental goals: How evolutionary psychology misconceives human rationality. In D. E. Over (Ed.), *Evolution and the psychology of thinking: The debate.* Psychological Press.
Stenning, K., & Van Lambalgen, M. (2008). *Human reasoning and cognitive science.* Cambridge, MA: MIT Press.
Stich, S. (1980). Could man be an irrational animal? Some notes on the epistemology of rationality. *Synthese, 64,* 115–135.
Stich, S. (1990). *The fragmentation of reason.* Cambridge, MA: MIT Press.
Sugiyama, L. S., Tooby, J., & Cosmides, L. (2002). Cross-cultural evidence of cognitive adaptations for social exchange among the Shiwiar of Ecuadorian Amazonia. *Proceedings of the National Academy of Sciences, 99,* 11537–11542.
Todd, P. M. (2001). Fast and frugal heuristics for environmentally bounded minds. In G. Gigerenzer & R. Selten (Eds.), *Bounded rationality: The adaptive toolbox* (pp. 51–70). Cambridge, MA: MIT Press.
Todd, P. M., Hertwig, R., & Hoffrage, U. (2005). Evolutionary cognitive psychology. In D. Buss (Ed.), *The handbook of evolutionary psychology* (pp. 776–802). Hoboken, NJ: Wiley.
Trivers, R. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology, 34,* 35–57.
Vollmer, G. (1975). Evolutionäre Erkenntnistheorie. Stuttgart: Hirzel.

The Central Role of Cognitive Search in Bounded and Adaptive Rationality

Wai-Tat Fu **Saturday 23 November, 11:00 AM**

How do humans or animals adapt to a new environment? After years of research, it is embarrassing how little we understand the underlying processes of this important form of adaptation: just look at how difficult it is to build a robot that learns to navigate in a new environment or to teach someone to master a second language.

It is amazing how seagulls and vultures have learned to be landfill scavengers in the last century and be able to sort through human garbage to dig out edible morsels. It seems obvious that the ability to adapt to new (or changing) environments goes beyond hardwired processes that exploit structures of the environment, but also relies on the ability to acquire new knowledge about the environment, in ways such that rational actions can be chosen to attain important goals. Some natural questions to ask are: What “should” a rational actor do to keep this balance of exploiting known structures and exploring for new knowledge about the environment? What should be the criteria to judge whether an actor is rational when adapting to a new environment?

The goal of this paper is to investigate the role of *cognitive search* in theories of bounded and adaptive rationality. Cognitive search can be defined as the broader set of processes that allow the actor to achieve a goal by sampling through the environment in face of uncertainties about the nature, location, or methods of entities that will be encountered in the environment (Fu, Hills, & Todd). The uncertainties involved in search imply that the actor’s estimates of the environmental structures are bounded by his internal representation derived from the sampling process. Based on this definition, cognitive search is necessary not only when the actor adapts to a new environment, but is essential for any probabilistic textured environment (Brunswik, 1943). Focusing on the role of cognitive search, adaptive rationality can be studied with respect to (1) the sampling process, (2) the formation and updating of the internal representations of the environment, and (3) the choice of actions that achieve the actor’s goals. The paper will discuss when and how adaptive rationality in one of more of these components may lead to behavior that may differ from certain normative standards of performance.

More generally, the paper will also argue that search is not just one of the cognitive processes that reflects adaptive rationality, but it is the central process that is essential for adaptive rationality. In fact, the literature in cognitive science and artificial intelligence has long established that search enables actions to be intelligent (Newell & Simon, 1976)—in the sense that much of the intelligence in cognitive computations can be derived from observing how the agent performs search, such as the ability to extract information from the task environment and that information to guide the search, avoiding wrong turns and dead ends. To a large extent, search serves as the process that allows the organism to interface with the environment to intelligently achieve its goals. Principles of adaptive rationality are important to understand when and how intelligent behavior emerges from the cognitive search process.

The paper will first provide a definition and overview of cognitive search, and how it is useful for understanding bounded and adaptive rationality. A brief review of the literature from the domains of artificial intelligence and cognitive psychology on how search is performed by machines and humans will be provided, with a special focus on its relation to adaptive rationality. Specific examples, mostly based on theories and models of information search (e.g., Fu, 2007; Fu & Gray, 2006; Fu & Pirolli, 2007; Pirolli & Card, 1999), will also be provided to illustrate how cognitive search is pivotal to human intelligence, and how it reflects adaptive rationality.

References

Brunswik, E. (1943). Organismic Achievement and Environmental Probability. *Psychological Review 50,* 255–272.

Fu, W.-T., Hill, T., & Todd, P. (to appear in 2014). Special issue on "Interfacing mind and environment: The central role of search in cognition". *Topics in Cognitive Science.*

Fu, W. T., & Gray, W. D. (2006). Suboptimal tradeoffs in information seeking. *Cognitive Psychology, 52,* 195–242

Fu, W. T., & Pirolli, P. (2007). SNIF-ACT: A model of information-seeking on the World Wide Web. *Human-Computer Interaction, 22,* 355–412.

Fu, W.-T. (2007), Adaptive Tradeoffs Between Exploration and Exploitation: A Rational-Ecological Approach. In W. D. Gray (Ed.), *Integrated Models of Cognitive Systems.* Oxford: Oxford University Press

Newell, A., & Simon, H. A. (1976). Computer science as empirical inquiry: symbols and search. *Commun. ACM, 19*(3), 113–126.

Pirolli, P., & Card, S. K. (1999). Information Foraging. *Psychological Review 106*(4), 643–675.

Strategy-Selection by Meta-Induction in Prediction and Decision Tasks under Changing Environments: Theoretical and Empirical Results

Gerhard Schurz
Paul Thorn
Christian Feldbacher

Saturday 23 November, 2:00 PM

Prediction methods produce predictions about unobserved events, in dependence on available information about the given environment. Decision methods can be understood as based on prediction methods predicting which action (among all actions that are available at a certain time) will have maximal payoff. We understand the *rationality* of prediction and decision methods in an *external* sense, as given by their success and efficiency. Their rationality has (at least) three dimensions: They have to produce (a) many true predictions or useful decisions with (b) low computational costs in (c) as many different environments as possible.

According to the paradigm of *universal rationality,* good reasoning strategies should be as general and environment-independent as possible, being successfully applicable to virtually all cognitive purposes in all types of environments. In philosophy, this paradigm was promoted in its "deductive" variant by critical rationalists (Popper 1935) and in its inductive variant by logical empiricists (Carnap and Jeffrey 1971). In *psychology,* the universalist paradigm is represented by the Turing model of cognition (e.g., Newell and Simon 1972), and in its inductive variant by, for example, universal learning theories based on behavioral conditioning (e.g., Bittermann 2000).

The paradigm of universal rationality has been subjected to serious criticism by advocates of the younger paradigm of *bounded and adaptive rationality,* who argue that good prediction and decision methods should be, and typically are, adapted to specific environments, being tailored to specific tasks for which they provide highly efficient solutions. In analytic philosophy, this paradigm is rather new. In psychology, this paradigm was defended by Simon (1982), who argued that all-purpose mechanisms are hopelessly inefficient. In a stronger form, this paradigm was defended by Gigerenzer et al. (1999, 2001) and Hertwig et al. (2011): their research program operates on the assumption that all successful prediction and decision methods are more or less local, and that simple cognitive heuristics are frequently more successful than slow general reasoning mechanisms.

In spite of their evolutionary plausibility, locally adapted prediction or decision methods have the disadvantage that their performance is highly dependent on the *environment* in which they are applied. However, biological organisms and especially humans frequently encounter situations of rapidly changing environments. In recognition of this fact, the present paper is devoted to the question of adaptive rationality under changing environments. Under these conditions, one needs strategies which select for a given environment a method which performs well in this environment. We call these strategies *meta-level methods,* in opposition to *object-level methods.* Meta-methods can only be successful if their success is at least partially independent from the environment—that is, holds for a class of environments which is as broad as possible. So in contrast to object-level methods, the applicability of meta-level methods should be highly general, without incurring a dramatic increase of cognitive complexity.

There is no reason why evolution, apart from having selected domain-specific methods, should not have also selected general-purpose meta-strategies, which make it possible for organisms to adapt to frequently changing environments. The surplus effect of combining locally adapted and general reasoning mechanisms has been emphasized in psychology within the *dual account of cognition* (cf. Over 2003). The present paper can be understood as developing a special version of the dual account of cognition, arguing for the superiority of a dual combination of general meta-level methods which are superimposed on a given set of available locally adapted object-level methods.

An outstanding class of meta-methods are meta-inductive methods, whose investigation is the topic of the paper. The meta-inductivist tracks the so-far observed performance of competing object-level methods, and on that basis aims to select or construct an optimal prediction method. For example, if it is true that in some environments laymen are better forecasters than experts in regard to the stock market (as reported in Gigerenzer et al. 1999, 60ff.), then the meta-inductivist will favor laymen over experts in these environments. It has been shown in previous work that meta-inductive strategies are optimal in a very broad range of environments, which means that if added to a given class of available object-level methods they can only improve the maximal success of these methods in the long run, even under the assumption of a changing environment in which the most successful object-level methods are constantly changing.

There exist different variants of meta-inductive methods. The simplest meta-inductive strategy is *Imitate-the-best,* which uses the so-far most successful method to predict future events. While *Imitate-the-best* always emulates only one existing (but possibly changing) prediction method, *Weigh-the-bests* employs a weighted average of the predictions of those strategies which have been observed to be most successful so far, where the weights are determined by the past success rates. Schurz (2008, 2009) has investigated meta-induction in the form of so-called prediction games. Roughly speaking, a prediction game consists of a sequence of events e_1, e_2,… coded by real (or binary) numbers, and a set of players P_1, P_2,… whose task at time or round n it is to deliver a prediction p_{n+1} of next event e_{n+1} (both represented by a real number in [0,1]), where the performance of the players is assessed by a linear or convex scoring function. Schurz (2012a) shows that the formal structure of prediction games can be applied to decision games, by understanding a decision for an action A_i as the prediction that the action A_i will have maximal payoff in the next round n+1. It was demonstrated that *Imitate-the-best* is only optimal in a restricted class of environments, in which the success rates of methods converge rapidly to constant frequency limits. The same result applies to the method *Take-the-best* (TTB) of Gigerenzer et al. (1999, ch. 2–4), which can be understood as a version Imitate-the-best in which the role of object-level methods is played by "cues." The most important theoretical result is the existence of a certain method of weighted meta-induction which is provably optimal in the long run in arbitrary environments (if applied to finitely many object-level methods), and whose short-run success has precisely calculable upper bounds which rapidly converge to zero when the number of rounds of the game becomes much larger than the number of competing object-level methods (Schurz 2008, based on theorems in Cesa-Bianchi and Lugosi 2006). There exist several variants of weighted meta-induction by which one may increase the success rate of meta-induction and reduces its short-run loss: for example by using different weighing functions, or by conditionalizing the success-rates of the object-level methods in which the meta-inductivist bases her prediction on a restricted past horizon or on certain types of environments.

In the first and *theoretical part* of the paper, we will present theoretical and simulation results about different variants of *Imitate-the-best* and *Weigh-the-bests* meta-induction. We show that the impressing success rate of Imitate-the-best which has been demonstrated by Gigerenzer et al. (1999, ch. 2–4) under the assumption of constant success rates (or "ecological validities") breaks down in certain "adversarial" situations of changing environments. Drawing on earlier results we demonstrate that *Weigh-the-best* is universally long-run optimal even in these adversarial situations. Most importantly, we show that if the success rates of object-level methods correlate strongly with certain types of environments which are recognizable by the meta-inductivist, the success of meta-induction can be significantly improved by conditionalizing the success-evaluation on these types of environments.

In the second and *empirical part* of the paper, we apply different variants of meta-induction in retrospect to real-life prediction competitions. Most importantly, we apply meta-induction to the results of the Monash University Probabilistic Footy Tipping Competition, a competition which involves predicting the results of the Australian Football League (AFL). This competition is now 19 years old and is one of the world's longest-running prediction competitions. It is maintained by David Dowe and Torsten Seemann at the Faculty of Information Technology at the Australian Monash University (see www.csse.monash.edu.au/~footy). We treated the sequence of matches of one AFL season as one prediction game. The competition is probabilistic, which means that the forecasters had to specify their subjective probability for the winning of one of the two teams of a given match. They were evaluated by a linear or logarithmic scoring rule (increasing inversely with the distance between the predicted probability value and the real event, which is 1 if the favored team wins, 0 if it looses, and 0.5 in the case of a draw). The probability estimates of all participants for each AFL season were retrieved from an internet-accessible database. The predictions of meta-inductive forecasters were then determined by applying the meta-inductive method to the real forecasters. This procedure enabled us to determine what the performance of the meta-inductive methods would have been had they participated in the AFL prediction tournament. The evaluation was achieved with the help of a computer program based on Thorn and Schurz (2012).

According to our knowledge, the performance of meta-induction in real-life situations of this sort has so far not been tested. The results were especially interesting, inasmuch in the given test situation the meta-inductive predictions were forced to operate in the difficult situation of a comparatively *short* prediction horizon with a *high* number of competing prediction methods—and it is known that the worst case performance of meta-inductive methods in such a situation is relatively poor. Surprisingly, the actual short-run losses were much smaller than the theoretically calculable worst case short-run loss. In many cases the meta-inductive predictions were even superior in the short run. We will report various further exciting results. For example, it turned out that by certain refinements (e.g., by assessing the performance of a method by only considering the accuracy of its most recent predictions), the performance of the meta-inductive forecaster can be significantly improved so that it became the best forecaster even in the short run. However, these refinements are not guaranteed to work in all environments, and there are no general mathematical theorems abut their long-run behavior. So at this point, the perspective of local adaptivity as the "ultima ratio" seems to return within the meta-level.

As a second empirical study we intend to apply meta-inductive method to different methods of predicting changes in stock prices (cf. Feldbacher 2012). Research relevant to this part of our paper is still in progress, and the results achieved within the next few months shall be included in the final paper.

References

Barkow, J. H., Cosmides, L., & Tooby, J. (Eds.). (1992). *The Adapted Mind. Evolutionary Psychology and the Generation of Culture.* New York: Oxford Univ. Press.

Bitterman, M. (2000). Cognitive Evolution. In C. M. Heyes & L. Huber (Eds.), 61–80.

Carnap, R., & Jeffrey, R. (1971). *Studies in Inductive Logic and Probability.* Berkeley, CA: Univ. of California Press.

Cesa-Bianchi, N., & Lugosi, G. (2006). *Prediction, Learning, and Games.* Cambridge, MA: Cambridge Univ. Press.

Feldbacher, C. (2012). „Meta-Induction and the Wisdom of Crowds". *Analyse und Kritik, 34*(2), 367–382.

Gigerenzer, G., Todd, P. M., & the ABC Research Group (Eds.). (1999). *Simple Heuristics that Make Us Smart.* Oxford: Oxford Univ. Press.

Gigerenzer, G., & Selten, R. (Eds.). (2001). *Bounded Rationality. The Adaptive Toolbox.* Cambridge, MA: MIT Press.

Hertwig, R., Hoffrage, U., & the ABC Research Group. (2011). *Simple Heuristics in a Social World.* New York: Oxford University Press.

Over, D. E. (2003). From Massive Modularity to Metarepresentation: the Evolution of Higher Cognition. In D. E. Over (Ed.), *Evolution and the Psychology of Thinking: The Debate* (pp. 121–144). Hove, UK: Psychology Press.

Schurz, G. (2008). The Meta-Inductivist's Winning Strategy in the Prediction Game: A New Ap- proach to Hume's Problem. *Philosophy of Science, 75,* 278–305.

Schurz, G. (2009). Meta-Induction and Social Epistemology: Computer Simulations of Prediction Games. *Episteme, 6,* 200–220.

Schurz, G. (2012a). Meta-Induction and the Problem of Fundamental Disagreement. In C. Jäger & W. Löffler (Eds.), *Epistemology: Contexts, Values, Disagreement* (pp. 343–354). Frankfurt a. M.: Ontos Verlag.

Schurz, G. (2012b). Meta-Induction in Epistemic Networks and Social Spread of Knowledge. *Episteme, 9*(2), 151–170.

Simon, H. A. (1982). *Models of Bounded Rationality* (Vol. 1, pp. 164-187), Cambridge, MA: MIT Press.

Thorn, P., & Schurz, G. (2012). Meta-Induction and the Wisdom of Crowds. *Analyse und Kritik, 34*(2), 339–366.

Rational Task Analysis

Hansjörg Neth
Chris R. Sims
Wayne D. Gray

Saturday 23 November, 4:00 PM

The rationality of minds and machines is bounded by and adapted to internal limits and external environments (Simon, 1955, 1956). With regards to the appropriate method to investigate bounded rationality, Newell and Simon's (1972)notorious scissors analogy called for "both an analysis of the structure of task environments and an analysis of the limits of rational adaptation to task requirements." (p. 55). But despite this early emphasis on the interactive and environmentally embedded nature of bounded rationality most empirical investigations of rational behavior still follow a pattern that is deeply rooted in the logic of experimental design. To assess the adaptive potential of cognition, researchers manipulate environmental variables and measure the extent to which (human or non-human) participants can cope with the specific contingencies that govern a particular condition.

A common finding of such studies presents advocates of rationality with a puzzle: Organisms frequently remain stuck in a behavioral pattern of stable suboptimal performance (Herrnstein, 1991; Fu & Gray, 2004)—that is, they exhibit insufficient adaptations to particular task environments. If trivial explanations (like insufficient instruction or motivation) can be excluded, stable suboptimal behavior reveals the limits of boundedly rational organisms.

While we agree that experimental attempts at mapping the bounds of rationality are intriguing, we caution that their conclusions are often premature. More specifically, exposures of the bounds of rationality often tend to be based on insufficient analyses and a biased perspective on experimental task environments. To prevent such premature claims in the future, we introduce the methodological approach of *rational task analysis* (RTA). By translating Newell and Simon's (1972) scissors analogy into a sequence of methodological steps, RTA helps to prevent errors in investigations of seemingly irrational behavior, such as premature assessments of irrationality.

To illustrate and promote our methodology, we present several case studies that show the application of RTA to particular tasks and task environments (Neth, Sims, Veksler, & Gray, 2004; Neth, Khemlani, & Gray, 2008; Sims, Neth, Jacobs, & Gray, 2013). In each case, a systematic re-analysis that involved a shift in perspective resulted in the revocation of an earlier diagnosis that seemed to expose some lack of rationality in experimental participants. Distilling the common elements that motivated these re-evaluations will allow us to define and explicate our methodological approach in general. We conclude by comparing RTA to related approaches (e.g., Anderson, 1990; Hertwig & Erev, 2009; Howes, Lewis, & Vera, 2009) and identifying findings and paradigms that seem promising candidates for RTA in the future.

References

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ, USA: Lawrence Erlbaum Associates.

Fu, W.-T., & Gray, W. D. (2004). Resolving the paradox of the active user: Stable suboptimal performance in interactive tasks. *Cognitive Science, 28*(6), 901–935.

Herrnstein, R. J. (1991). Experiments on stable suboptimality in individual behavior. *The American Economic Review, 81*(2), 360–364.

Hertwig, R., & Erev, I. (2009). The description-experience gap in risky choice. *Trends in Cognitive Sciences, 13*(12), 517–523.

Howes, A., Lewis, R. L., & Vera, A. (2009). Rational adaptation under task and processing constraints: Implications for testing theories of cognition and action. *Psychological Review, 116*(4), 717.
Neth, H., Khemlani, S. S., & Gray, W. D. (2008). Feedback design for the control of a dynamic multitasking system: Dissociating outcome feedback from control feedback. *Human Factors, 50*(4), 643–651.
Neth, H., Sims, C. R., Veksler, V. D., & Gray, W. D. (2004). You can't play straight TRACS and win: Memory updates in a dynamic task environment. In *Proceedings of the 26th annual meeting of the Cognitive Science Society* (pp. 1017–1022). Hillsdale, NJ: Lawrence Erlbaum.
Newell, A., & Simon, H.-A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice Hall.
Simon, H. A. (1955). A behavioral model of rational choice. T*he Quarterly Journal of Economics, 69*(1), 99–118.
Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63*(2), 129–138.
Sims, C. R., Neth, H., Jacobs, R. A., & Gray, W. D. (2013). Melioration as rational choice: Sequential decision making in uncertain environments. *Psychological Review, 120*(1), 139–154.

Fast and Frugal Learning

Sarah Wellen
David Danks

Sunday 24 November, 10:00 AM

The goal of the bounded/adaptive rationality research programme is to study how cognitively limited individuals make decisions in an uncertain world. The largest (by far) strand of this programme argues that these individuals succeed by using "fast and frugal" heuristics that take advantage of evolved cognitive capacities/limitations and the structure of the environment to make successful decisions. A variety of heuristics have been proposed for search, inference, judgment, and decision making, but almost no attention has been devoted to understanding the adaptive nature of *learning.* This lacuna is particularly surprising since appropriate representations are critical for successful decision making. In this paper, we argue that many types of learning can and should be studied from the adaptive rationality perspective. Of course, achieving this goal requires that we make precise what it means for a learning heuristic to "succeed" in a given environment. We argue that this depends critically on the pragmatic and epistemic goals of the learner.

Learning can be broadly understood as a process by which people form and revise mental representations for later use. Learning processes have been largely ignored in the study of bounded rationality, as most researchers have simply assumed the existence of the relevant mental representations for inference or decision making. There have been some exceptions. Anderson & Schooler's (1991) rational model of memory can be understood as a model of how memories (i.e., one type of representation) are modified over time based on usefulness. Learning has also been studied implicitly: most bounded rationality models of a cognitive task assume that people have some representation that must presumably have been learned. For example, research on the Take the Best (TTB) heuristic initially assumed that people represent (and so must learn) cue validities. In the course of investigating TTB, however, it was discovered that people instead learn about the "success value" of a cue (Newell, Rakow, Weston, & Shanks, 2004; Rakow, Hinvest, Jackson, & Palmer, 2004). This example points to the interaction between learning and decision making: bounded rationality requires not only that inference and decision making be "fast and frugal" given some representations, but also that those representations be acquired through "fast and frugal" learning.

We outline a framework for understanding the adaptive rationality of a learning process: does it allow bounded individuals to achieve their goals with limited and uncertain evidence? Boundedly rational learning should employ heuristics with many of the same qualitative features as those found in search, inference, and decision making. In particular, we argue that they should (i) be fast and frugal (ii) take advantage of evolved capacities/limitations, and (iii) allow the learner to achieve their goals in their natural environment. Criterion (iii) points to a key challenge for boundedly rational learning, as the goals of learning are often quite unclear. Learned representations do not immediately engage with the environment (in contrast with, say, decisions or actions), and so there is not necessarily a natural "measuring stick" for whether a learned representation is successful by some measure.

One natural measure is that learning is successful just when it allows the learner to acquire true and/or justified beliefs about the world. That is, perhaps learning is always for epistemic goals. We argue, however, that this is not actually the proper response from the bounded rationality perspective. People sometimes do adopt purely epistemic goals, but they also often learn in order to achieve pragmatic goals like bringing about a desired outcome or succeeding on a particular task. Success at these pragmatic goals need not require complete, or sometimes even accurate, representations of the world; the truth is not always best, or even necessary, for a task. More generally, the success of learning is determined by whether the acquired representations allow the agent to achieve her goals—whether epistemic or pragmatic—in inference or decision making.

This general argument that learning should be goal-dependent is arguably uninteresting in the abstract; a potentially more interesting issue is how this conclusion changes the ways that we study and understanding learning. We thus turn to consider three different cases of learning in which the bounded rationality approach yields novel predictions. First, some learning heuristics that have been thought to be irrational (or at least, non-normative) are arguably adaptively rational for particular epistemic goals. For example, people use order of observations as a cue to causal structure (Lagnado, Waldmann, Hagmayer, & Sloman, 2007), even when they know that observational order may not follow the true sequence of events (Lagnado & Sloman, 2006; Sloman & Lagnado, 2005). This "temporal sequence" heuristic can be formalized and defended as adaptively rational for our natural environment. Similarly, Monte Carlo approximations of Bayesian inference (e.g., Shi, Griffiths, Feldman, & Sanborn, 2010; Bonawitz, Denison, Chen, Gopnik, & Griffiths, 2011) provide examples of processes that satisfy purely epistemic goals by approximating (in a boundedly rational way) complex functions by using simple evolved capacities (e.g., exemplar-based reasoning).

Second, many learning heuristics can be defended as adaptively rational for pragmatic goals. For example, suppose the learner wishes to succeed at a particular, finitely describable task (e.g., choose between actions a_1 or a_2 based on some feature set F) in a relatively stable environment. The constraints of bounded rationality imply that a learning mechanism is rational if it allows the learner to make the correct decision with minimal computational effort and minimal evidence. A natural implication for this situation is that the learner *ought* to learn only about distinctions that have practical consequences for decision making. This general constraint will typically be relevant in determining the very space of possibilities over which learning takes place. All else being equal, it is harder to learn when one has more hypotheses or possibilities (e.g., generative models in a Bayesian framework, regions in feature space in categorization models, parameters in some model, etc.), and so a frugal learner ought to make the minimum number of distinctions necessary for achieving success at the tasks she expects to encounter. This general observation already helps to make sense of some empirical data that has been taken to show irrational behavior.

The third case study applies this general moral to the particular case of causal learning. For given evidence, a boundedly rational agent ought to attend to different features depending on whether she wishes to (a) learn the true causal structure; (b) make predictions from observations; or (c) choose interventions to bring about an outcome. (a) is an epistemic goal while (b) and (c) are both pragmatic goals, and so it is unsurprising that they have different informational requirements. Goal (a) requires one to learn the full causal and probabilistic structure. Goal (b) requires information only about conditional probabilities, so a frugal learner can make fewer (or no) distinctions between causal structures. Goal (c) requires partial information about causal structure (e.g., which variables are direct or indirect causes of the target), but less information about conditional probabilities (e.g., *only* which variables will have the greatest impact, not a quantitative measure of that aspect). We report the results of ongoing empirical work investigating the ways in which people sometimes learn different representations depending on their goals.

We conclude by considering challenges for applying the adaptive rationality approach to learning. In particular, it is usually possible in the laboratory to provide a precise description of the representations necessary to succeed at a well-defined goal (and thus determine the value of a learning heuristic within this context). In real life, there is typically significant uncertainty about both future environments and also future pragmatic/epistemic goals. Learning about every feature of the world that could potentially be relevant sometime in the future is, of course, completely irrational for bounded agents. There is, however, limited work—experimental, observational, or theoretical—on the challenge of predicting one's own future needs and the ways in which those predictions feed back into learning. Nonetheless, our theoretical arguments and experimental data show the importance of considering learning processes from the perspective of bounded rationality.

References

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological science, 2*(6), 396–408.

Bonawitz, E., Denison, S., Chen, A., Gopnik, A., & Griffiths, T. L. (2011). A Simple Sequential Algorithm for Approximating Bayesian Inference. In L. Carlson, C. Holscher, & T. Shiple (Eds.), *Proceedings of the 33rd Annual conference of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Lagnado, D. A., & Sloman, S. A. (2006). Time as a guide to cause. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(3), 451.

Lagnado, D. A., Waldmann, M. R., Hagmayer, Y., & Sloman, S. A. (2007). Beyond Covariation. In A. Gopnik & L. Schultz (Eds.), *Causal learning: Psychology, philosophy, and computation* (p. 154). USA: Oxford University Press.

Newell, B. R., Rakow, T., Weston, N. J., & Shanks, D. R. (2004). Search strategies in decision making: The success of "success". *Journal of Behavioral Decision Making, 17*(2), 117–137.

Rakow, T., Hinvest, N., Jackson, E., & Palmer, M. (2004). Simple heuristics from the Adaptive Toolbox: Can we perform the requisite learning? *Thinking & reasoning, 10*(1), 1–29.

Shi, L., Griffiths, T. L., Feldman, N. H., & Sanborn, A. N. (2010). Exemplar models as a mechanism for performing Bayesian inference. *Psychonomic bulletin & review, 17*(4), 443–464.

Sloman, S., & Lagnado, D. A. (2005). Do we 'do'. *Cognitive Science, 29,* 5–39.

The Benefits of an Inclusive Approach to the Study of Rationality

Michael Bishop
Heather Cipoletti

Sunday 24 November, 12:00 PM

What is it to reason in a rational manner? This is a question that both philosophers and psychologists wrestle with. It is central to both disciplines. The traditional philosophical approach to building a theory of rationality supposes that the only substantive evidence a theory must answer to is our pretheoretical judgments (our "intuitions") about rationality (Cohen 1981). Over the past two decades or so, this "armchair" approach has come under increasing scrutiny (Stich 1990; Weinberg, et al. 2001; Kornblith 2002; Bishop & Trout 2005). If we abandon the traditional approach, what should we put in its place? We will argue for an "inclusive" approach to the study of rationality. This approach begins with the idea that a good theory of rationality will provide a unified, coherent account of two lines of evidence:

1. Philosophers' epistemological theorizing.
2. Psychologists' empirical research into how reasoning works and how it can be improved.

On the one hand, our theories of rationality cannot completely ignore our commonsense intuitions. Rationality cannot turn out to be a banana. But on the other hand, our theories need to do much more than just answer to our commonsense intuitions. We have no good reason to think our intuitions on these matters are so pristine that they deserve to be the only evidence a theory of rationality must answer to. A good theory of rationality must be tied down to our best findings about how to go about teaching people to reason better about the world.

We will sketch a theory that drops out of the inclusive approach to the study of rationality. We don't need to start from scratch since this theory will share the basic contours of Bishop & Trout's *Strategic Reliabilism* (2005, 2008). However, the theory is an improvement on Strategic Reliabilism insofar as it allows us to account more coherently for a wider array of evidence. It resolves—or at least helps clarify—certain outstanding problems that arise in philosophical and psychological research on rationality.

References

Bishop, M., & Trout, J. D. (2005). *Epistemology and the psychology of human judgment.* Oxford: Oxford University Press.

Bishop, M., & Trout, J. D. (2008). Strategic reliabilism: A naturalistic approach to epistemology. *Philosophy Compass, 3* (5), 1049–1065.

Cohen, L. J. (1981). Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences 4* (3), 317–331.

Kornblith, H. (2002). *Knowledge and its place in nature.* Cambridge, MA: MIT Press, 2002.

Stich, S. (1990). T*he fragmentation of reason: preface to a pragmatic theory of cognitive evaluation.* Cambridge, MA: Bradford Books/MIT Press.

Weinberg, J., Nichols, S., & Stich, S. (2001). Normativity and epistemic intuitions. *Philosophical Topics, 29*(1-2), 429–60.