Minds & Machines (2016) 26:125–148
DOI 10.1007/s11023-015-9368-8

Rational Task Analysis: A Methodology to Benchmark Bounded Rationality

Hansjörg Neth[1,2] · Chris R. Sims[3] · Wayne D. Gray[4]

Received: 15 May 2013 / Accepted: 16 April 2015 / Published online: 7 May 2015

Abstract How can we study bounded rationality? We answer this question by proposing *rational task analysis* (RTA)—a systematic approach that prevents experimental researchers from drawing premature conclusions regarding the (ir-)rationality of agents. RTA is a methodology and perspective that is anchored in the notion of bounded rationality and aids in the unbiased interpretation of results and the design of more conclusive experimental paradigms. RTA focuses on concrete tasks as the primary interface between agents and environments and requires explicating essential task elements, specifying rational norms, and bracketing the range of possible performance, before contrasting various benchmarks with actual performance. After describing RTA's core components we illustrate its use in three case studies that examine human memory updating, multitasking behavior, and melioration. We discuss RTA's characteristic elements and limitations by comparing it to related approaches. We conclude that RTA provides a useful tool to

✉ Hansjörg Neth
neth@mpib-berlin.mpg.de;
http://neth.de

Chris R. Sims
chris.sims@drexel.edu;
http://www.pages.drexel.edu/∼crs346

Wayne D. Gray
grayw@rpi.edu;
http://www.rpi.edu/∼grayw/

[1] Center for Adaptive Behavior and Cognition (ABC), Max Planck Institute for Human Development, Berlin, Germany

[2] Social Psychology and Decision Sciences, University of Konstanz, Konstanz, Germany

[3] Drexel University, Philadelphia, PA, USA

[4] Rensselaer Polytechnic Institute, Troy, NY, USA

render the study of bounded rationality more transparent and less prone to theoretical confusion.

Keywords Bounded rationality · Benchmarking · Optimality · Task environment · Rational analysis · Ecological rationality

Just as a scissors cannot cut paper without two blades, a theory of thinking and problem solving cannot predict behavior unless it encompasses both an analysis of the structure of task environments and an analysis of the limits of rational adaptation to task requirements.
(Newell and Simon 1972, p. 55)

Introduction

Assessments of rationality play a central role in the analysis of minds and machines. However, the utility of such an endeavor requires not just a norm of rationality but a method of measuring adherence to that norm. Herbert Simon's pioneering insight that human cognition is both bounded by and adapted to its environment redefined the yardstick by which behavior ought to be measured. His notion of *bounded rationality* (Simon 1955, 1956, 1990) conveys that agents perform tasks with limited information, computational capacity, and time, and implies that rationality is a joint function of an agent's cognitive capacity and environmental resources. Nevertheless, many inquiries into the nature of rationality lack a careful analysis of the environment or of the interplay between internal and external constraints.

Regarding the rationality of the human species, a vast amount of research portrays people as blundering simpletons more reminiscent of Homer Simpson than of Homo sapiens. Fraught with "general misconceptions" (Ross and Nisbett 1991, p. 86) and suffering from "heuristic biases" (Ferguson 2008, p. 346) that cause "severe and systematic errors" (Tversky and Kahneman 1974, p. 1124) we must strive to curb our primitive "animal spirits" (Akerlof and Shiller 2010) and hope for turning out "predictably irrational" (Ariely 2008). As a consequence, scientists have accumulated large arsenals of chronic flaws and biases,[1] concluded that "mental illusions should be considered the rule rather than the exception" (Thaler 1994, p. 4), and equated the task of mapping bounded rationality with "exploring the systematic biases that separate the beliefs that people have and the choices they make from the optimal beliefs and choices assumed in rational-agent models" (Kahneman 2003, p. 1449). The verdict that people routinely violate rational norms has led to a lively theoretical debate, in which researchers' "inordinate fondness for errors" (Krueger and Funder 2004, p. 317) has been criticized as a rhetoric of irrationality (Lopes 1991) and a distorted view of cognitive illusions (Gigerenzer 1991). Unfortunately, some key methodological implications of the notion of bounded rationality have been lost in the trenches of these "rationality wars" (Samuels et al. 2002).

[1] Krueger and Funder (2004, Table 1, p. 317) provide a "partial list" of 42 errors of judgment, and http://en.wikipedia.org/wiki/List_of_cognitive_biases (retrieved on Dec. 22, 2014) collects over 180 cognitive biases, many of which can be re-interpreted as smart adaptations (Gigerenzer 2004).

Newell and Simon's (1972, p. 55) 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". Yet despite this early emphasis on the interactive and environmentally embedded nature of bounded rationality, most empirical investigations of rational behavior (Burns 2002; Herrnstein and Vaughan 1980; Shakeri and Funk 2007) follow the traditional logic of experimental research design. To assess the adaptiveness of cognition, researchers manipulate the properties of task environments and evaluate the extent to which organisms cope with the manipulated contingencies. Any substantial deviation of behavior from the experimenter's conception of optimal task performance is then diagnosed as an anomaly or irrationality. A common finding among studies adopting this approach is that organisms exhibit insufficient adaptations to particular task environments and remain stuck in a behavioral pattern of stable suboptimal performance (Fu and Gray 2004; Herrnstein 1991). If trivial explanations (like insufficient instruction or motivation) can be excluded, stable suboptimal behavior is believed to reveal the limits of boundedly rational organisms.

Although laboratory-based attempts at mapping the bounds of rationality are an invaluable source of empirical data, we think that the conclusions drawn from these data are often premature. More specifically, sightings of the bounds of rationality are frequently based on a biased perspective on experimental task environments. This bias conflates the experimenter's understanding of the task—and hence optimal task performance—with the experimental subject's limited knowledge of the environment. The result of this *experimenter bias* is an increase in Type-1 errors in the investigation of bounded rationality, i.e., the premature acceptance of the claim that behavior deviates from some rational norm. Our intent in documenting the experimenter bias is constructive rather than destructive. In Homer's *The Odyssey*, the hero Odysseus is able to avoid the call of the Sirens by recognizing his weakness and binding himself to the mast of his ship. Similarly, by documenting the dangers of experimental bias in the assessment of rationality, we hope to steer investigators away from rocky shoals in the interpretation of experimental results.

Our contribution to this end is the methodological approach of *rational task analysis* (RTA). Rather than suggesting a general conclusion about the rationality of behavior or integrating competing theories of human cognition, RTA provides a tool to aid experimental design and the unbiased interpretation of research results. By translating Newell and Simon's (1972) scissors analogy into a set of methodological principles, RTA offers a safeguard against misadventures in the mapping of bounded rationality. Hence, RTA is not a new theory, but rather a methodology—a proposal of best practices for the interpretation of results and experimental research design regarding the study of bounded rationality.

In this article, we first present an overview of the core components of RTA. By explicating key aspects of the task and bracketing the range of plausible human performance, RTA provides a methodology to benchmark bounded rationality and measure the impact of environmental interventions on behavior. To be applicable to a wide array of research questions and domains, RTA is expressed as a checklist of caveats that should be considered before drawing conclusions about an organism's rationality. We then present three studies that catalyzed our development of RTA

and illustrate its applicability across different tasks and task environments. In each case, an initial claim of irrational behavior turns out to be the result of an experimenter's biased or incomplete understanding of the task environment. RTA revises the premature diagnosis and provides novel insights in the performance of rational agents or the properties of the task environment. We conclude by comparing our proposal to related approaches and suggest that RTA—as a flexible methodology that can accommodate different theoretical perspectives—is an essential tool to explore and enlighten the nature of bounded rationality.

The Core Components of RTA

Rational task analysis (RTA) is a methodology and perspective that is anchored in the notion of bounded rationality and "encompasses both an analysis of the structure of task environments and an analysis of the limits of rational adaptation to task requirements" (Newell and Simon 1972, p. 55). By translating the metaphor of Simon's scissors into a set of methodological principles, RTA is a tool for conducting and interpreting rationality research and provides an answer to the question: How can we study bounded rationality? RTA's main purpose is to prevent premature conclusions regarding the rationality or irrationality of agents performing specific tasks. As a methodology, RTA can be applied to examine existing experiments and corresponding results, as well as to modify existing or design new research paradigms. To map the bounds of rational behavior, any instance of RTA focuses on a concrete task as the primary interface between rational agents and the environment.

Box 1 summarizes the core components of our approach. The process begins whenever a research question arises regarding the rationality of an agent performing

Box 1 Overview of the core components involved in conducting a rational task analysis (RTA)

1. State the *research question* and *rational behavior* to be addressed
2. Define the *task*, and key features of the *agent* and *task environment*:
   (a) What is the *goal* of the task?
   (b) What is the agent's *motivation* to perform the task?
   (c) Which *resources* or *constraints* enable or limit performance?
   (d) Which *criterion* is used to evaluate task performance?
3. Bracket the range of *possible performances* by mathematical modeling or agent-based simulations. Relevant *benchmarks* to be determined are:
   (a) One or more lower bounds of *baseline* performance
   (b) One or more upper bounds of *optimal* performance
   (c) Optional benchmarks to measure the performance of *specific strategies*
4. Collect *data* and contrast actual performance with the benchmarks
5. Consider *interventions* to the task environment and repeat Items 2–4
6. Conclude or iterate

some specific task (Item 1). In practice, however, research projects rarely start from scratch, but instead begin with some notable finding or previous conclusion. As most complex tasks pose many challenges to cognitive agents and can be studied from several perspectives, explicitly stating the specific question and behavior addressed helps preventing misattributions and overly specific or general conclusions.

A key stance of our approach is that the *task* (Item 2) is the most useful unit of aggregation to study bounded rationality and defines the interface between agent and environment, i.e., the edge along which Simon's scissors aim to cut. Rather than rushing to collect new empirical data, RTA first explicates the details of the studied task by adopting a functional notion of the task environment (Gray et al. 2006) which is jointly determined by features of the task (e.g., its goal and constraints), the agent (e.g., its goal, motivation, internal capacities, and constraints), and the environment (e.g., the performance criterion and availability of resources). In addition to focusing on a task's definition and objective, a functional perspective assumes that internal and external constraints define and shape the task in an adaptive and interactive fashion. For instance, the task of chopping down a tree is functionally quite different from chopping wood for the fireplace and depends crucially on features of the axe (e.g., its weight distribution, a sharp or dull blade) and of the agent (an axeman's physique and experience).[2]

The notion of benchmarking (Item 3) is a central and fundamental component of RTA. The use of baseline and optimal benchmarks to map the range of possible performances is a generalization of the *bracketing heuristic* (Gray and Boehm-Davis 2000; Kieras and Meyer 2000), which originated in the context of computational cognitive modeling and used a slowest- and fastest-reasonable model to estimate expected processing times and guide the design of realistic models. Instead of focusing on latencies, RTA requires lower and upper benchmarks to delimit the range of possible performances.

RTA allows for different kinds of benchmarks, provided that they are explicated and justified. Typical lower bounds for task performance consist in distributing actions randomly or uniformly over all options available in an environment. Specifying an upper bound for performance requires explicating a norm of optimality. The concept of multiple optima is no oxymoron when considering that agents can be endowed with different types of background knowledge. We distinguish between three different kinds of knowledge and corresponding norms of optimality: With *certain knowledge*, all relevant aspects of the task environment are known to the agent. By contrast, an agent behaves *under risk* when future outcomes are probabilistic, but the relevant probabilities are known or can be estimated (Knight 1921). Lastly, agents act *under uncertainty* whenever the possible consequences of choices are unknown or it is difficult or impossible to assign probabilities to an exhaustive list of outcomes. As different assumptions about an agent's knowledge yield different types of optimality, RTA's flexibility regarding the selection of a rational norm is no flaw, but an essential feature. For instance, the plurality of optima under different types of knowledge allows for different

[2] See Scriven (1991, p. 346) for an elaboration of this example.

perspectives on a task by experimenters (who analyze and design tasks under certainty or risk) versus participants (who mostly act under risk or under uncertainty). If lower and upper bounds are well-defined, the realistic range of actual performance falls within the range of possible performance. Including additional benchmarks for specific strategies (e.g., simple heuristics or algorithms with theoretically-motivated constraints) allows gauging the range of reasonable performance.

The fact that empirical data collection (Item 4) occurs relatively late in RTA does not diminish its importance. Instead, all preceding elements prepared the ground to properly evaluate agents' actual performance. When contrasting observed behavior with benchmarks, we must bear in mind that deviations from lower and upper bounds are to be expected, particularly when aggregating measures over tasks or individuals. Whereas observing agents perform a task at baseline level may indicate some lack of incentive or task-relevant information, meeting a norm of optimality is not a reasonable demand on rational behavior. In fact, reserving the credit of rationality for demonstrable optimality and merely measuring deviations from this standard would enshrine rationality as a null hypothesis that can only be falsified (Krueger and Funder 2004, p. 318). In addition to reiterating our need for multiple benchmarks (Item 3), the untenable status of an isolated norm of rationality has two important consequences: First, any positive verdict for an agent's rationality requires specifying an alternative hypothesis (Nickerson 2000) or a meaningful difference from optimality (similar to a statistical effect size). Second, suboptimal behavior can be considered rational if it falls within some tolerated range of optimality or if agents have good reasons for falling short of an optimal benchmark (e.g., when the behavioral costs of actions outweigh their benefits, or when limited capacities or resources constrain performance).

Introducing interventions to the task environment or changing the task by considering potential moderators (Item 5) is an optional component of RTA. It is not uncommon that comparisons between benchmarks and actual behavior yield new hypotheses that suggest how performance could be boosted by changing some aspect of the agent, task, or task environment (e.g., by altering instructions, providing additional incentives, or highlighting task-relevant information). Typically, any substantial modification of the task requires an iterative cycle through Items 2–4 until a conclusion can be reached.

Just as our approach does not prescribe the specifics of benchmarks or potential interventions, RTA offers no predetermined result (e.g., guarantee that some particular behavior will be found to be rational or irrational) and has no predefined completion criterion (Item 6). As with any other methodology, the interpretation of research findings remain the responsibility of the researcher. Thus, RTA structures the shape of an argument and ends when an investigator following its principles is confident to have accumulated enough evidence to reach a conclusion regarding the rationality of an agent performing a specific task.

Overall, the set of core components summarized in Box 1 must not be viewed as a fixed sequence of steps or an invariable recipe. Instead, they provide a collection of best practices or checklist of caveats, which should be examined prior to judging an agent's rationality. As a methodology, RTA is best thought of as a swiss-army

knife that provides a flexible set of tools, any one of which may turn out useful or essential, but not all are always necessary to get a particular job done. To prevent arbitrariness, RTA is anchored in the notion of bounded rationality and rests on the core principles of examining concrete tasks, explicating assumptions and norms, and contrasting actual performance with a range of precisely specified benchmarks. In the following, the flexible yet systematic use of RTA will be illustrated by three case studies, which catalyzed the development of our approach and demonstrate its applicability across different tasks and psychological domains.

Three Case Studies

TRACS: A Baseline Bias in Dynamic Decision Environments?

TRACS™ is a 'Tool for Research on Adaptive Cognitive Strategies' in the form of an experimental card game that was designed to investigate dynamic decision making under risk (Burns 2001). Playing the game consists in making a series of choices in which a player turns over one of two hidden cards to match the color of a third. To maximize the number of matches over a total of 11 turns, a player should track the changing odds of cards as the deck is being depleted. For example, at the beginning of each game, the number of red and blue cards in the deck is equal. As

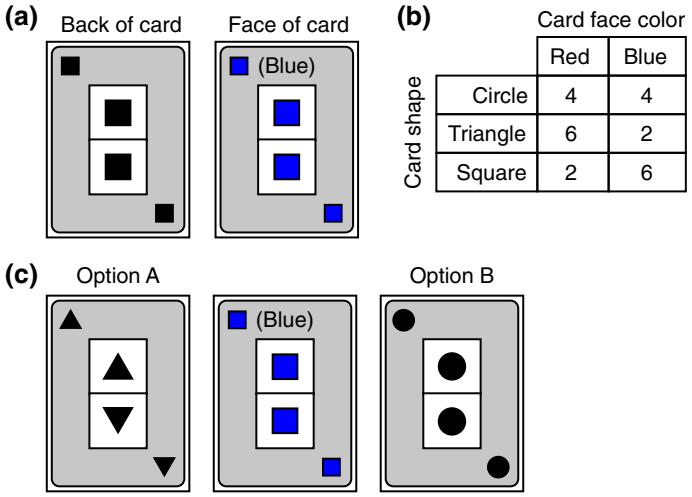

Fig. 1 The game of TRACS. **a** Each card in the deck has a black colored back, showing a shape (*square*, *circle*, or *triangle*), and a front showing both shape and color (*red* or *blue*). **b** Frequency of each card type in the deck (containing 24 cards). **c** On each of 11 turns, three cards are dealt. The center card is dealt face up, showing its color. A player must choose to turn over either the left or the right card. The goal is to choose the card that is more likely to match the color of the center card. In this example, Option B is more likely to match the color of the center card, since *circle* cards are more likely to be *blue* than *triangle* cards. After every choice the chosen card is turned over, revealing its color, and both face-up cards are removed from the deck. As the game progresses, the frequencies of card types change

cards are removed from the deck the odds of uncovering a red versus blue card change. Based on an experiment and agent-based simulations, Burns (2002) reported a *baseline bias*: Rather than updating their memory with each turn according to the principles of Bayesian rationality, players seemed to base their choices on the initial distribution of cards in the deck. Figure 1 illustrates the mechanics of the game.

Although a limited capacity for dynamic memory updates is in line with previous research (Venturino 1997), an insufficient sensitivity to recent changes seems a serious threat to an organism's survival in dynamic environments. For example, in natural habitats, foraging animals should update their memory as food resources become depleted in one area and more prevalent in another. By contrast, the results obtained in TRACS suggest that human memory remains 'stuck in the past'. This anomaly of a baseline bias not only presents a challenge for any adaptive account of memory (Anderson and Schooler 1991) but also contradicts previous claims that people suffer from the opposite fallacy of base rate neglect (Tversky and Kahneman 1974).

Our analysis of TRACS (Neth et al. 2004) began with the question: How useful are memory updates in this task environment? If a failure to update memory had no significant impact on performance, a baseline bias might be irrational—in the sense of deviating from the normative ideal of Bayesian belief updating—while being simultaneously harmless. If the cognitive costs of memory updates exceeded their benefits, the observed insensitivity to changing frequencies could be a case of boundedly rational behavior. To examine this issue, we conducted a series of simulations that benchmarked the range of possible performance by four cognitive agents. A random agent provides a lower bound on performance by simply guessing on each choice. A slightly more sophisticated baseline agent chooses cards based on their initial distribution in the deck, but does not update its beliefs during the game. To explore the upper limits of performance, an ideal memory-updating agent accurately tracks all changing odds and chooses cards accordingly. Finally, an omniscient agent acts as if it was endowed with X-ray vision. Although this faculty dispenses with the need for memory altogether, it may still not yield a match on every turn, as some turns have no winning card. Note that the latter agents define two distinct optima: Whereas the memory-updating agent achieves the best possible performance of human players under risk, the omniscient agent marks a purely theoretic ceiling of the game by choosing cards under certainty.

A typical beginning player of TRACS succeeds in matching colors on 5–7 out of 11 turns and shows little improvement thereafter. To our surprise, our agents exhibit a similarly narrow range of performance, scoring 5.2 points for the random agent and 8.2 points for the omniscient agent (on average across 10,000 simulated games). Crucially, the mean performance of the ideal memory-updating agent barely exceeds that of the static baseline agent (6.8 vs. 6.6 points, respectively). Moreover, any noticeable benefit of the ideal memory-updater over a static baseline agent only occurs on the last five of 11 turns and presupposes perfect memory for the first half of the game.

We concluded that the original game provides insufficient incentives for adopting an effortful memory update strategy. This verdict allows for the possibility that

players could remember more *if* the game provided additional motivation to do so. To examine this, we constructed a variant of the game. Our modified version TRACS* (Neth et al. 2004) looks identical to and obeys the same rules as the original game, but stacks the deck so that a dynamic updating strategy outperforms a baseline strategy by a larger margin. As predicted, the behavior of humans playing the modified game provided a more optimistic view of their capacity for memory updates. Although players did not achieve the level of performance of an ideal memory-update agent, they outperformed the baseline agent when memory mattered for performance. In addition to the behavioral evidence of task performance beyond the baseline level, human players explicitly reported odds that corresponded more closely to the actual odds than to the baseline odds.

Although our study made its point, it also had clear limitations. Our findings provided an existence proof for a basic capacity for dynamic memory updates, but did not investigate the details of our players' motivations or potential. If human memory updates reflect a cost-benefit tradeoff, and hence a boundedly rational strategy, it will be important to measure and quantify the costs on memory imposed by the task in future work. Thus, while our study exemplified the methodology of RTA, elucidating the underlying mechanisms would require a study of additional boundary conditions.

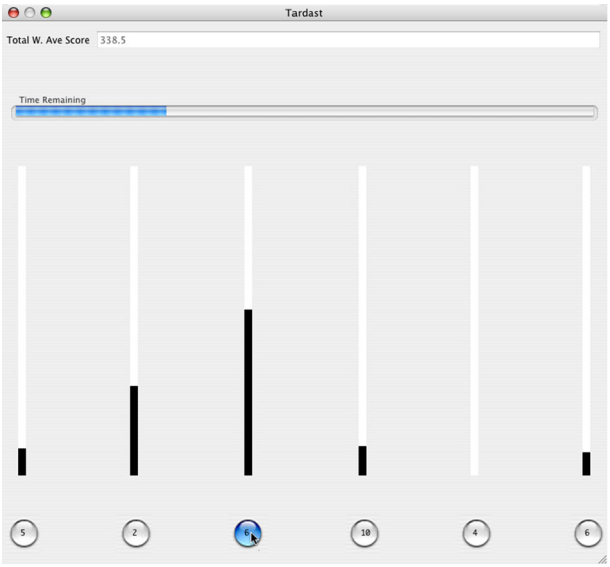

Fig. 2 Illustration of a Tardast scenario with six concurrent tasks. Each *vertical bar* represents a task, and the height of the *black portion* indicates its current satisfaction level *SL*. Pressing one of the *buttons* underneath a *bar* increases the *SL* of the corresponding task until it reaches its maximum of 100 %. The *button* values (5, 2, 6, etc.) specify each task's weight *W*. On every system cycle, the *W*s and all current *SL*s are integrated linearly to update a numerical feedback score (*top left*). The *horizontal progress bar* indicates the remaining scenario time

Tardast: Suboptimal Resource Allocation in Juggling Multiple Tasks?

Tardast is named after the Persian term for 'juggler' and provides an abstract and interactive framework to study human multitasking (Shakeri 2003; Shakeri and Funk 2007). The analogy to a juggler's feat of simultaneously spinning plates on vertical poles captures the resource allocation problem that lies at the core of any multitasking situation: Due to inherent limits of perceptual, cognitive, and action resources, organisms need to negotiate tradeoffs when facing several tasks at once. While working on any particular task, an operator or juggler needs to monitor the state of alternatives and the overall system to decide when to switch to another task. Although performance in Tardast is measured by a linear function of all task states over time, Shakeri (2003) proves that finding an optimal time-to-task allocation constitutes an NP-hard problem. Figure 2 illustrates an experiment designed to study cognitive resource allocation, in which a human operator must manage six concurrent tasks.

In experimental and simulation studies, Shakeri and Funk (2007) contrasted human performance with the near-optimal performance of a machine-learning algorithm and found human operators to be lacking in comparison. Human shortcomings were attributed to suboptimal time-to-task allocations and poor strategic task management. As the complexity of the system seemed to exceed human resource limitations, operators were judged to fail at adequately balancing and prioritizing tasks.

Our first study using the Tardast paradigm replicated and extended the original phenomenon of operators' stable suboptimal performance (Neth et al. 2006). By controlling for learning effects and comparing human performance not just to optimality, but also to artificial agents that provided benchmarks for baseline and simple heuristic performances, our study yielded two insights: First, different Tardast scenarios are clearly not of equal difficulty. Specifically, whenever some scenarios generally afford higher scores than others, absolute scores cannot be compared across scenarios. As the upper and lower benchmarks across scenarios varied in parallel to human performance scores, our bracketing strategy revealed that performance differences between scenarios were largely explained by environmental differences.[3]

A second result painted human performance in an even more sobering light than the findings by Shakeri and Funk (2007). For many scenarios, human operators barely performed above baseline and were demonstrably suboptimal, not only relative to a normative ideal, but also when compared to a simple heuristic strategy that operators could easily have implemented. Taken together, these results left us with a puzzle: Why were human operators so utterly dependent on environmental characteristics and fail to discover or implement simple strategies to boost their performance?

[3] Behavioral patterns that closely mirror the shape of task environments are reminiscent of Simon's ant-on-the-beach analogy: "The apparent complexity of our behavior over time is largely a reflection of the complexity of the environment in which we find ourselves" (Simon 1996, p. 53).

A second study was motivated by an intervention to the task environment and identified lack of control on the part of Tardast operators as a piece of the puzzle (Neth et al. 2008). Cognitive processes can be constrained by limited processing resources or by data limits (Norman and Bobrow 1975). When searching for potential data limits in Tardast, we realized that the original system provided *outcome* feedback, as the displayed score reflected an operator's overall achievement over the entire scenario. We contrasted this increasingly inert numeric feedback score with the notion of *control* feedback, which provides moment-to-moment guidance for action selection. Hypothesizing that perfect outcome feedback may still provide suboptimal control feedback, we replaced the original score with an alternative one that dynamically reflects current system quality (i.e., metaphorically, showing a ship's current speed, rather than its overall distance traveled so far). An experiment and further simulations showed that human operators allocated their attention and actions more adaptively when equipped with the new control feedback mechanism. Although the original outcome feedback provided the very measure by which performance was evaluated, control feedback—by virtue of being more responsive to local system changes—facilitated superior performance outcomes. This demonstrated that the poor performance of Tardast operators was not just caused by human capacity limits but also by data limits of the original feedback score. Again, our analysis and study of a slightly modified task environment succeeded in boosting human performance beyond the levels previously observed. But although operator performance with control feedback reached the level of simple heuristics, it still remained suboptimal.

The Harvard Game: Myopia and Melioration in Choice Under Uncertainty?

Psychology and behavioral economics are rife with empirical demonstrations of human choice violating the norms of rational choice theory. The research program on heuristics and biases (Kahneman et al. 1982; Tversky and Kahneman 1974) has cataloged a wide array of such examples: People inappropriately frame decision problems, evaluate consequences with respect to an inappropriate reference point, or fail to apply the rules of probability theory in reasoning about risks. In all these cases, the consequence of irrational choice leaves the experimental participant worse off. Thus, while people might state a preference for more money, their actual choices seem to leave them with less.

Richard Herrnstein and colleagues suggested that another factor may be at work in suboptimal choice. As an alternative source of suboptimality they investigated a phenomenon called *melioration* (Herrnstein and Vaughan 1980), which refers to the process of choosing an alternative that has the highest immediate utility. On the surface, melioration does not seem such a bad choice strategy. However, meliorating behavior ignores the fact that immediate actions can often have negative consequences for future utility. For example, the decision to eat an unhealthy meal today may have small, but negative consequences for future health and happiness. Thus, choosing to maximize local gains might entail serious long-term pains.

Much of the research behind melioration theory was built upon studies of animal choice (Herrnstein 1982; Herrnstein and Vaughan 1980; Vaughan 1981). In these experiments, animal subjects (typically pigeons) faced instrumental choice tasks between two alternative sources of food. Summarizing the results of a typical experiment, Herrnstein writes "In the experiment as a whole, the pigeons earned food at a lower rate than they would have by allocating choices randomly to the two alternatives, let alone what they could have earned as food reinforcement maximizers" (Herrnstein 1990, p. 362). So much for pigeon rationality.

But are humans just as bird-brained? In numerous studies, human participants have also been shown to meliorate by systematically favoring locally rewarding options over global utility maximization, even when melioration is the worst possible decision strategy for the task (Rachlin and Laibson 1997). This persistent tendency to meliorate has widely been viewed as an explanatory factor for phenomena as diverse as natural selection (Dawkins 1999), stable suboptimal performance in touch-typing (Yechiam et al. 2003), addiction (Herrnstein and Prelec 1992; Heyman and Dunn 2002), delinquency (Wilson and Herrnstein 1985), as well as impulsivity and lack of self-control (Herrnstein 1981). Consequently, Herrnstein (1990, p. 218) concluded that "behavior is generically suboptimal, though still orderly, and that optimality is the exception rather than the rule".

Figure 3 illustrates the central aspects of a task environment—known as the *Harvard game*—that has previously been shown to induce meliorating behavior in humans (Rachlin and Laibson 1997). In this experimental paradigm, participants are informed that they must make a number of n choices (e.g., $n = 800$) between two alternatives (labeled as Options A and B) and instructed to maximize their overall gains. After each choice, they may or may not receive a monetary reward, and then face the same choice again on the next trial. Crucially, the probability of obtaining a reward on each trial and for each of the alternatives must be learned from experience and depends on participants' past history of choices. In Fig. 3, the probability of reward is plotted for each of the two alternatives as a function of the participant's preference for Option A on the previous ten choices. Option B *always* has a higher probability of yielding a reward, but the more often it is chosen, the worse both options become (i.e., reward probabilities for both options are a decreasing function of preference for Option B). In this environment, the strategy that maximizes total winnings is to always choose the option with locally worse prospects (Option A). Humans, like pigeons, tend to meliorate in this task and demonstrate a stable bias towards Option B.

Based on our successful interventions in TRACS and Tardast we first conducted two experiments that explored the potential of more global feedback mechanisms in helping people to maximize rewards (Neth et al. 2005, 2006). When these attempts failed, we subjected the Harvard game to an RTA (Sims et al. 2013). This involved a crucial change in perspective: Previous experiments using this paradigm had neglected to appreciate that the experimenter's knowledge of the task is very different from the participant's view on it. From the perspective of the experimenter, meliorating behavior in the Harvard game exemplifies irrational choice under risk. However, for an experimental participant, the relationship between actions and consequences is uncertain and must be learned from

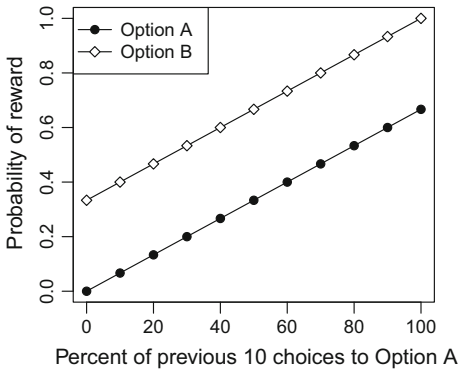

Fig. 3 Reward contingencies of the Harvard game, a decision environment designed to discriminate between global maximization (by consistently choosing *Option A*) and melioration (by preferring the locally superior *Option B*)

experience. Thus, the key question in the Harvard game is: What should a rational participant be *expected* to know or learn about this environment, given a finite amount of experience with the task? To answer this question, we designed an optimal Bayesian agent and set it loose on the Harvard game.[4] To our own surprise, even an unboundedly rational agent would be expected to meliorate, and persist in this supposedly irrational strategy for thousands of trials. Figure 4 plots the predicted reward rate for exclusive preference to each of the two alternatives according to the Bayesian learning agent, as a function of the amount of its experience with the task. The Bayesian learning agent demonstrates that a preference for melioration should rationally persist for nearly twenty thousand trials. Not only did our RTA unmask the rationality of melioration, but it also was able to explain relatively subtle aspects of human performance in the experiment. For example, each participant in the experiment experienced a slightly different task, as a consequence of making different choices and observing different outcomes. The Bayesian learning model was able to predict how this individuated experience should influence future choices. Importantly, participants who observed greater evidence indicative of the globally optimal strategy in fact showed a smaller bias towards melioration.

In summary, our RTA of the Harvard game (Sims et al. 2013) questions its suitability to support far-reaching conclusions about the fundamental irrationality of human choice in tasks with indirect and delayed consequences. By developing a Bayesian model of the learning problem faced by individuals in uncertain decision environments, we demonstrated that an unbiased learner would adopt melioration as the optimal response strategy for maximizing long-term gain. Focusing on the nature of the task as perceived by the participant, rather than as assumed by the experimenter, suggests that many documented cases of melioration should not be interpreted as irrational choice, but can be understood as globally optimal choice under uncertainty.

[4] This Bayesian agent formalized the learning task as one of inferring a posterior distribution over the relevant history window of environmental states, a function that maps each choice history onto one of a discrete number of states, and the probability of obtaining a reward for choosing either option in each possible state of the environment (see Sims et al. 2013, p. 143 ff., for details).

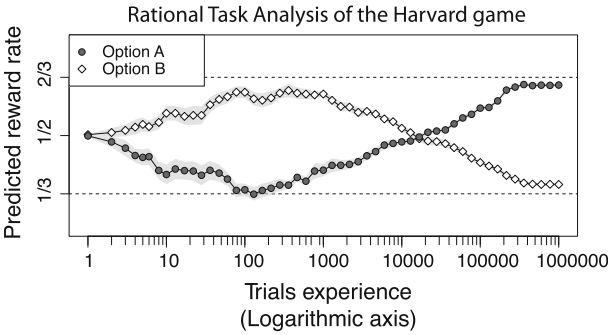

Fig. 4 Predicted reward rates of the rational learner model (for the maximizing *Option A* and meliorating *Option B*) in the Harvard game

Discussion

After introducing and illustrating the methodology of RTA, we now compare our case studies to discuss their similarities and differences and highlight some essential and accidental elements of our approach. Briefly mentioning some related approaches will uncover many common themes, but also reveal important limitations of RTA.

Similarities and Differences Between our Case Studies

Our three case studies exemplified the elements characterizing our general approach (cf. Box 1 on p. 4): First, we took someone else's research question and diagnosis of irrational behavior—in the form of a baseline bias, suboptimal multitasking performance, or a persistent bias towards melioration—as the beginning of a research program, rather than its end. Second, our RTA began with a careful examination of a concrete task and environment that seemed to have elicited some behavioral anomaly. To capture the functional aspects of the task, we specified key features of cognitive agents and their task environments (e.g., the current goal, performance criterion, as well as task-relevant resources and constraints). Third, we created computational or formal models of minimal and optimal task performance to bracket the range of possible performance. The purpose of these simulations was not the design of cognitive models with a maximum of naturalistic constraints, but the provision of benchmarks for actual performance. Fourth, we collected and compared empirical performance data with our simulated benchmarks. Fifth, we considered changes to some task environments (TRACS and Tardast) and hypothesized that they would facilitate performance. Implementing and assessing the effects of these interventions required additional simulations and data collections. Sixth, and finally, we concluded that the task environments were ill-suited to answer the original research questions and that claims regarding agents' alleged irrationality had to be revoked or qualified.

Despite these similarities, our studies also differed in important aspects. Their first and most striking difference is that they addressed dissimilar phenomena and psychological domains, which serves as an indication of RTA's broad generality. Second, although all examples crucially relied on simulated performance benchmarks, their details, goals, and consequences varied between studies. In TRACS, a narrow gap between a lower and two upper benchmarks (for optimal performance under certainty vs. risk) revealed that human performance was not as biased as originally believed. In Tardast, we never expected people to perform optimally, but were concerned about their failure to match the performance of a simple heuristic. In both task environments, the range of possible performance would remain unknown without suitable benchmarks. By contrast, determining the lower and upper bounds on performance in the Harvard game seemed trivial, as they follow directly from its definition (see Fig. 3). In this context, our demonstration of a difference between optimal performance under risk versus uncertainty made a theoretical contribution that questioned the traditional interpretation of results based on this paradigm. Third, our analysis of TRACS and Tardast both suggested interventions that improved performance, but led to different interpretations. As our modifications of TRACS were invisible to agents, they demonstrated an ability to outperform a pure baseline strategy that could not be detected in the original game. In Tardast, our implementation of a more responsive feedback score showed that the poor performance in the original paradigm was partly due to a lack of control feedback. Fourth, and finally, our case studies reached different conclusions. Although all three original claims of irrationality were premature, agents still exhibited limitations in our revised versions of TRACS and Tardast. By contrast, our Bayesian model of the Harvard game showed that agents were consistent with the behavior of an ideal rational learner.

Accidental Versus Essential Features of RTA

Taken together, these comparisons allow us to distinguish between the accidental and essential features of our approach. For instance, it was merely accidental that all three of our reported case studies responded to someone else's work and resulted in the revision or qualification of an earlier claims regarding agents' alleged irrationality. Rather than merely being reactive by suggesting *post hoc* critiques of existing results and experiments, all core components of RTA can be used to explore uncharted research territory and construct and evaluate new paradigms. Similarly, RTA has no pre-determined result or implicit agenda to demonstrate the universal rationality of human cognition. Thus, the similar conclusions of our studies were not caused by any aspect of RTA, but due to chance or our selection of examples. By merely subscribing to the basic tenets of bounded rationality, RTA remains epistemically neutral regarding the rationality of any particular behavior under investigation. In fact, providing no guarantees for showing an agent's rationality is a precondition for rendering any such verdict informative and convincing.

The most obvious essential feature of RTA is its commitment to the notion of bounded rationality and the systematic use of methodological core components. By

examining concrete tasks, explicating norms and benchmarks that delimit the range of possible performance before contrasting them with actual performance, RTA makes behavior more measurable and prevents premature conclusions regarding the rationality of agents.

Defining RTA as a methodology and perspective highlights two essential aspects: First, a methodology is no ideology. Beyond conceptualizing rationality as a joint function of agents and task environments RTA does not subscribe to a particular norm or theory of rationality. This openness is intentional, as it allows researchers from different backgrounds to adopt our approach. The frameworks of Bayesian rationality (Oaksford and Chater 2007), ecological rationality (Gigerenzer et al. 1999; Todd et al. 2012), computational rationality (Lewis et al. 2014), and social rationality (Hertwig et al. 2013), each offer ready-to-hand definitions of optimal performance in a given task. By contrast, RTA's non-committal stance allows for a kind of meta-experimentation, by examining the changes in conclusions obtained when plugging-in different yardsticks for measuring rationality.

Second, adopting RTA entails a profound change in perspective. Traditional research on rationality views and evaluates an agent's behavior from the perspective of an omnipotent experimenter, who is not only the judge to issue verdicts of irrationality, but also defines the standard and designs the test by which an agent's rationality is being measured. If a systematic and substantial deviation from a rational norm is found, three possible ways to defend the agent's rationality are (a) choosing a different norm, (b) criticizing the way in which a norm is being measured, and (c) designing a different test. All of these routes have been explored in the past (for examples, see Birnbaum's 1983 and Koehler's 1996, critiques of base-rate neglect; the debate between Kahneman and Tversky 1996 and Gigerenzer 1996 on heuristics; or the re-analysis of the hot-hand fallacy by Burns 2004). RTA does not favor one of these alternatives, but makes a contribution by turning the task environment into an object of study and showing it from the perspective of a rational agent. As experimental environments are a matter of design, different designs can yield different results. Crucially, experimenters are often blind to or myopic about the appearance of their designs from the agent's viewpoint. For instance, our analysis in Sims et al. (2013) was based on the premise that agents in probabilistic environments behave and learn on the basis of limited experience, i.e., inhabit an uncertain world, rather than the risky one designed by its creator. By simulating optimal performance from the agent's perspective, RTA helps experimenters to overcome their blind spot and to determine a task's suitability for measuring an explicated norm of rationality.[5]

Another essential feature of our approach is its flexibility. We mentioned in Sect. 2 that RTA does not prescribe a fixed sequence of steps, is open to multiple norms, and can be understood as a flexible set of tools or collection of best practices. Our comparisons between our case studies have shown that their courses and

[5] Similar shifts of perspective are reported in the literature on *decision by sampling* (Fiedler and Juslin 2006; Stewart et al. 2006). The consequences of presenting risk-related information in different representational formats are explored in studies on the *description-experience gap* (Hertwig et al. 2004; Hertwig and Erev 2009). Both paradigms provide strong additional arguments for the adoption of a subject-based perspective when conducting research and interpreting experimental results.

conclusions were neither planned nor obvious when we began our investigations, but evolved during our analysis. Although RTA's potential to ask and flexibly respond to new questions is a valuable asset, it must not become arbitrary. Fortunately, RTA's flexibility is constrained by two protective factors—its commitment to the presence of several core components (see Sect. 2) and its constant emphasis on explication and justification. If two researchers or theoretic perspectives reached different conclusions regarding the interpretation of some finding, RTA would at least expose the source of—and hopefully help to resolve—their conflict.

Related Approaches and Limitations of RTA

Beyond RTA, the notion of bounded rationality (Simon 1955, 1956, 1990) has inspired a wealth of paradigms and theoretical approaches. Today, the frameworks of rational analysis (Anderson 1990), computational rationality (Lewis et al. 2014), heuristics and biases (Kahneman et al. 1982; Tversky and Kahneman 1974), ecological rationality (Gigerenzer et al. 1999; Todd et al. 2012), and social rationality (Hertwig et al. 2013) lay claims to Simon's heritage and enrich and elaborate his original ideas. Rather than being planned strategically, our approach evolved by addressing practical issues raised by our case studies and investigations of immediate interactive behavior (Neth et al. 2007). RTA developed to fill a methodological niche in the space defined by Simon's work, but has hardly been conceived in a theoretical vacuum. Comparing it to related and recent approaches that share a similar vision reveals some family resemblances, but also important limitations of RTA.

Rational Analysis

Anderson's (1990) *rational analysis* (RA) of cognition is our most immediate source of inspiration. To study cognition independently from its biological implementation, RA assumes that it is adapted to evolutionary important environments in an optimal fashion. Given an agent's goals, a formal model of the environment, and minimal assumptions about cognitive constraints, an optimal behavioral function can be derived and compared with empirical data. In case of a correspondence, RA explains cognition by assuming a functional view of the agent, a formal model of the environment, and a process of optimization. For instance, many effects of memory can be understood as optimal solutions to basic information-retrieval tasks (Anderson and Milson 1989).

Despite some overlap between RA's steps and RTA's core components (cf. Anderson 1990, p. 29, with Box 1, p. 4), there are substantial differences between both approaches. For instance, RA's notion of the environment is much wider than RTA's notion of the task. Whereas RA aims for formal models that capture environmental regularities in general, RTA examines concrete tasks (like TRACS, Tardast, or the Harvard game). RTA's smaller scale and scope can be seen as a limitation, but comes with gains in tractability. In fact, RA's goal of deriving an

optimal behavioral function may often be infeasible (Sanborn et al. 2010) whereas RTA's optimal benchmarks for concrete tasks are more readily obtained. Similarly, RTA's more modest focus on specific tasks allows a nuanced incorporation of mechanistic and task-specific constraints, whereas it is difficult to see how such moderators or interventions could be considered in RA's more abstract notion of environments. Both approaches also assign different weights and roles to the two blades of Simon's scissors. RA emphasizes the environmental blade and uses it as an input to derive insights about cognitive mechanisms. By contrast, RTA not only studies simpler and more specific scissors, but assumes a balanced interplay between both blades.[6] Finally, RA and RTA assign different roles to optimality. In RA, rationality and an evolutionary process of optimization are assumed as premises. By contrast, an explicated notion of optimality enters RTA as an upper bound on performance, whereas the rationality of some behavior serves as a dependent variable and possible conclusion. Thus, RTA is not just RA in a nutshell.

Computational Rationality

Two recent proposals concerned with the use of formal analysis to define and quantify rational behavior within specified tasks are *computational rationality* (CR) (Lewis et al. 2014) and its precursor *cognitively bounded rational analysis* (CBRA) (Howes et al. 2009). In the tradition of computational cognitive modeling, naturalistic constraints on cognition, perception, or motor control are typically implemented within a cognitive architecture (Newell 1990; Meyer and Kieras 1997; Anderson et al. 2004). CR interprets rational behavior as the solution to an optimal program problem that subjects both environmental and mechanistic constraints to a process of utility maximization. Although this can yield different definitions of rationality as a variation of environmental constraints, the primary focus of CR is on mechanistic constraints. Consequently, a key contribution of CR/CBRA is to enable rigorous testing of the assumptions embedded within alternative cognitive architectures.

Again, the premises and ambitions of our approach are different. RTA does not assume an optimization process on part of the organism and is not intended as a model evaluation or selection tool, but as a tool to aid and shape experimental design. Rather than testing alternative models of cognition, RTA helps developing tasks that allow to measure the bounds of rationality, and drawing unbiased conclusions from such tasks.

Heuristics and Ecological Rationality

Studying tasks of judgment and decision-making under risk and uncertainty, the research programs of *heuristics and biases* (Tversky and Kahneman 1974; Kahneman et al. 1982) and *fast-and-frugal heuristics* (Gigerenzer et al. 1999, 2011) agree that humans routinely rely on heuristics, but disagree about their

[6] RA's relative neglect of agent-based constraints was also responsible for Simon's skepticism towards this framework (Simon 1991).

evaluation as either involving inevitable trade-offs and systematic errors (Tversky and Kahneman 1974) or as adaptive tools yielding accurate and robust results under uncertainty (Gigerenzer et al. 1999, Neth and Gigerenzer 2015). RTA's nearest neighbor in this context is the notion of *ecological rationality* (ER), which involves a research program that explicitly "investigates the fit between the two blades of Simon's scissors" (Todd et al. 2012, p. 15).[7]

Within ER, rationality is understood as a matter of adaptive fit between a strategy, the environment, and the evolved or acquired capacities of agents. Examining the interplay of this triad to evaluate the degree of fit is a challenging process that overlaps substantially with RTA. Nevertheless, ER is both more specific and more general than our approach. Being rooted in the research tradition of judgment and decision-making, the strategies considered by ER typically target choice tasks and are implemented as process models of heuristics that mimic psychological mechanisms. By contrast, the strategies used as RTA's performance benchmarks address any type of task and allow for a wide range of models, including mathematical abstractions and models with strong architectural constraints. The more general nature of ER is evident in its focus on different types of environments. Whereas RTA studies and explicates the boundary conditions of specific task environments, ER discusses the structure of environments in more generic, often statistical terms (e.g., see Gigerenzer and Brighton 2009; Pleskac and Hertwig 2014). In addition, ER pursues a theoretical agenda that includes "a general vision of rationality" (Todd et al. 2012, p. 14) and questions traditional norms of rationality (e.g., logical consistency) by adopting adaptive success in real-world environments (i.e., efficient and effective solutions) as its main criterion.[8] By contrast, RTA contains no such theoretical commitment. Again, this can be viewed as a limitation, but provides benefits in flexibility. For instance, our critique of the Harvard game gained persuasive power by showing that melioration can be optimal even when adopting the traditional rational norm of an ideal Bayesian learner (Sims et al. 2013).

Other Threats to Validity

Despite its merits, RTA is no remedy against all experimental maladies and misconceptions. Its methodological nature implies not only that it promotes no new theory of cognition, but also includes no theory of the environment. This latter limitation distinguishes RTA from methodological approaches like *representative design* (RD) (Brunswik 1955, 1956), which is based on the theoretical framework of *probabilistic functionalism* (Brunswik 1943). RD extends the principle of representative sampling from subjects to stimuli to ensure that experimental conditions preserve the properties of natural environments.[9] As our approach

[7] See Todd and Gigerenzer (2001), for a comparison of Simon's scissors with the alternative metaphors of Shepard's mirror and Brunswik's lens.

[8] See the related notions of "achievement" and "correspondence" (Hammond and Stewart 2001).

[9] A volume edited by Hammond and Stewart (2001) provides an overview of Brunswik's essential contributions.

addresses the internal validity of rational arguments, it is equally applicable to study—in Brunswik's terms—"a mere homunculus of the laboratory", "a bearded lady at the fringes of reality", and "an ecological normal" (Brunswik 1955, p. 204).[10] Similarly, RTA provides no guarantee that subjects understand a task as intended by the experimenter. The experimenter bias examined by RTA involves a knowledge-gap between subject and designer regarding experimental contingencies. By contrast, a violation of construct validity due to a mismatch of goals (e.g., viewing a one-shot game as an instance of repeated choice, or a competitive situation as one calling for altruism) would be conceptualized as two different tasks. The fact that RTA was not developed to address these issues does not render it futile. Instead, researchers concerned about external validity and incongruent task constructs will be just as eager to embrace RTA to exclude additional sources of error.

Conclusion

> *Solving a problem simply means representing it so as to make the solution transparent.*
>
> (Simon 1996, p. 132)

This article asked the question: How can we study bounded rationality? The traditional answer to this question calls for controlled experiments that allow detecting the causes and conditions of rational behavior. Rather than demanding a radical departure from this practice, we point out that it easily leads to premature conclusions. As a potential remedy, we propose RTA—a methodology and perspective that analyzes specific tasks as the primary interface between agents and environments. Clearly, RTA is more demanding than the common practice of proclaiming an anomaly or bias whenever detecting another instance of stable suboptimal performance. But when aiming for firm and enduring conclusions this additional effort is indispensable and worthwhile. As RTA's key ingredients of explication, justification and benchmarking are not new, our main innovation lies in the systematic integration of existing parts in a general and powerful research methodology. Yet widespread familiarity with RTA's core components does not imply that it is simple or trivial. Not only does our approach differ from today's standard practice, but our case studies have shown that it can yield surprising results. Thus, we suggest that—as a package—RTA makes a contribution that deserves both a label and an audience.

In the above quote, Simon reminds us that a problem's transparent representation *is* its solution. RTA facilitates the study of bounded rationality precisely by rendering our arguments for and against the rationality of agents more transparent. A more widespread use of RTA would supply individual researchers with a useful tool to protect themselves from experimenter bias and to design experiments that

[10] A review of representative design and its impact on judgment and decision-making research is provided by Dhami et al. (2004).

allow for more robust and informative conclusions. On a more general level, adopting RTA provides a service to colleagues and the scientific community. By enabling a clearer communication of research results, RTA promises to reduce the theoretical confusion caused by implicit assumptions and inconclusive findings, and thus promote new insights into the nature of bounded rationality.

Acknowledgments We thank the attendants of the workshop on *Finding Foundations for Bounded and Adaptive Rationalityl* (taking place on Nov. 22–24, 2013, and organized by Ralph Hertwig, Arthur Paul Pedersen, and Renata Suter) as well as two anonymous reviewers for helpful feedback and suggestions.

References

Akerlof, G. A., & Shiller, R. J. (2010). *Animal spirits: How human psychology drives the economy, and why it matters for global capitalism*. Princeton, NJ: Princeton University Press.

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

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036–1060.

Anderson, J. R., & Milson, R. (1989). Human memory: An adaptive perspective. *Psychological Review*, *96*(4), 703–719.

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

Ariely, D. (2008). *Predictably irrational: The hidden forces that shape our decisions*. New York, NY: Harper Collins.

Birnbaum, M. H. (1983). Base rates in Bayesian inference: Signal detection analysis of the cab problem. *The American Journal of Psychology*, *96*(1), 85–94.

Brunswik, E. (1943). Organismic achievement and environmental probability. *Psychological Review*, *50*(3), 255–272.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*(3), 193–217.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley, CA: University of California Press.

Burns, B. D. (2004). Heuristics as beliefs and as behaviors: The adaptiveness of the "hot hand". *Cognitive Psychology*, *48*(3), 295–331.

Burns, K. (2001). TRACS: A tool for research on adaptive cognitive strategies: The game of confidence and consequence. http://www.tracsgame.com. Accessed 1 May 2015.

Burns, K. (2002). On straight TRACS: A baseline bias from mental models. In *Proceedings of the 24th Annual Meeting of the Cognitive Science Society* (pp. 154–159). Lawrence Erlbaum, Hillsdale, NJ.

Dawkins, R. (1999). *The extended phenotype: The long reach of the gene, revised edn*. Oxford, UK: Oxford University Press.

Dhami, M. K., Hertwig, R., & Hoffrage, U. (2004). The role of representative design in an ecological approach to cognition. *Psychological Bulletin*, *130*(6), 959–988.

Ferguson, N. (2008). *The ascent of money: A financial history of the world*. London, UK: Penguin.

Fiedler, K., & Juslin, P. (Eds.). (2006). *Information sampling and adaptive cognition*. New York, NY: Cambridge University Press.

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.

Gigerenzer, G. (1991). How to make cognitive illusions disappear: Beyond heuristics and biases. *European Review of Social Psychology*, *2*(1), 83–115.

Gigerenzer, G. (1996). On narrow norms and vague heuristics: A reply to Kahneman and Tversky. *Psychological Review*, *103*, 592–596.

Gigerenzer, G. (2004). The irrationality paradox. *Behavioral and Brain Sciences*, *27*(3), 336–338.

Gigerenzer, G., & Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Science*, *1*(1), 107–143.

Gigerenzer, G., Hertwig, R., & Pachur, T. (Eds.). (2011). *Heuristics: The foundations of adaptive behavior*. New York, NY: Oxford University Press.

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

Gray, W. D., & Boehm-Davis, D. A. (2000). Milliseconds matter: An introduction to microstrategies and to their use in describing and predicting interactive behavior. *Journal of Experimental Psychology: Applied*, *6*(4), 322–335.

Gray, W. D., Neth, H., & Schoelles, M. J. (2006). The functional task environment. In A. F. Kramer, D. A. Wiegman, & A. Kirlik (Eds.), *Attention: From theory to practice* (pp. 100–118). New York, NY: Oxford University Press.

Hammond, K. R., & Stewart, T. R. (2001). *The essential Brunswik: Beginnings, explications, applications*. New York, NY: Oxford University Press.

Herrnstein, R. J. (1981). Self-control as response strength. In C. M. Bradshaw, E. Szabadi, & C. F. Lowe (Eds.), *Recent developments in the quantification of steady-state operant behavior* (pp. 3–20). Amsterdam, NL: Elsevier.

Herrnstein, R. J. (1982). Melioration as behavioral dynamism. *Quantitative Analyses of Behavior*, *2*, 433–458.

Herrnstein, R. J. (1990). Behavior, reinforcement and utility. *Psychological Science*, *1*(4), 217–224.

Herrnstein, R. J. (1990). Rational choice theory. *American Psychologist*, *45*(3), 356–367.

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

Herrnstein, R. J., & Prelec, D. (1992). A theory of addiction. *Choice over time*, 331–360.

Herrnstein, R. J., & Vaughan, W, Jr. (1980). Melioration and behavioral allocation. In J. E. R. Staddon (Ed.), *Limits to action: The allocation of individual behavior* (pp. 143–176). New York, NY: Academic Press.

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science*, *15*(8), 534–539.

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

Hertwig, R., Hoffrage, U., & the ABC Research Group (2013). *Simple heuristics in a social world*. New York, NY: Oxford University Press.

Heyman, G. M., & Dunn, B. (2002). Decision biases and persistent illicit drug use: An experimental study of distributed choice and addiction. *Drug and Alcohol Dependence*, *67*(2), 193–203.

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.

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *American Economic Review*, *93*(5), 1449–1475.

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

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review*, *103*(3), 582–591.

Kieras, D. E., & Meyer, D. E. (2000). The role of cognitive task analysis in the application of predictive models of human performance. In J. M. C. Schraagen, S. E. Chipman, & V. L. Shalin (Eds.), *Cognitive task analysis* (pp. 237–260). Mahwah, NJ: Lawrence Erlbaum.

Knight, F. H. (1921). *Risk, uncertainty and profit*. Chicago, IL: University of Chicago Press.

Koehler, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and Brain Sciences*, *19*(01), 1–17.

Krueger, J. I., & Funder, D. C. (2004). Towards a balanced social psychology: Causes, consequences, and cures for the problem-seeking approach to social behavior and cognition. *Behavioral and Brain Sciences*, *27*(3), 313–327.

Lewis, R. L., Howes, A., & Singh, S. (2014). Computational rationality: Linking mechanism and behavior through bounded utility maximization. *Topics in Cognitive Science*, *6*(2), 279–311.

Lopes, L. L. (1991). The rhetoric of irrationality. *Theory and Psychology*, *1*(1), 65–82.

Meyer, D. E., & Kieras, D. E. (1997). A computational theory of executive cognitive processes and multiple-task performance. I: Basic mechanisms. *Psychological Review*, *104*(1), 3–65.

Neth, H., Carlson, R. A., Gray, W. D., Kirlik, A., Kirsh, D., & Payne, S. J. (2007). Immediate interactive behavior: A symposium on embodied and embedded cognition. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th Annual Meeting of the Cognitive Science Society* (pp. 33–34). Cognitive Science Society, Austin, TX.

Neth, H., & Gigerenzer, G. (2015). Heuristics: Tools for an uncertain world. In R. Scott & S. Kosslyn (Eds.), *Emerging trends in the social and behavioral sciences*. New York, NY: Wiley Online Library.

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., Khemlani, S. S., Oppermann, B., & Gray, W. D. (2006). Juggling multiple tasks: A rational analysis of multitasking in a synthetic task environment. In *Proceedings of the Human Factors and Ergonomics Society*, vol. 50, (pp. 1142–1146). Sage, San Francisco, CA.

Neth, H., Sims, C. R., & Gray, W. D. (2005). Melioration despite more information: The role of feedback frequency in stable suboptimal performance. In *Proceedings of the Human Factors and Ergonomics Society*, vol. 49, (pp. 357–361). Sage, Orlanco, FL.

Neth, H., Sims, C. R., & Gray, W. D. (2006). Melioration dominates maximization: Stable suboptimal performance despite global feedback. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th Annual Meeting of the Cognitive Science Society* (pp. 627–632). Lawrence Erlbaum, Hillsdale, NJ.

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 K. D. Forbus, D. Gentner & T. Regier (Eds.), *Proceedings of the 26th Annual Meeting of the Cognitive Science Society* (pp. 1017–1022). Lawrence Erlbaum, Hillsdale, NJ.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice Hall.

Nickerson, R. S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods*, *5*, 241–301.

Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology*, *7*(1), 44–64.

Oaksford, M., & Chater, N. (2007). *Bayesian rationality: The probabilistic approach to human reasoning*. New York, NY: Oxford University Press.

Pleskac, T. J., & Hertwig, R. (2014). Ecologically rational choice and the structure of the environment. *Journal of Experimental Psychology: General*, *143*(5), 2000–2019.

Rachlin, H., & Laibson, D. I. (Eds.). (1997). *The matching law: Papers on psychology and economics by Richard Herrnstein*. New York, NY: Russell Sage Foundation.

Ross, L., & Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology*. New York, NY: McGraw-Hill.

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, NY: Oxford University Press.

Sanborn, A. N., Griffiths, T. L., & Navarro, D. J. (2010). Rational approximations to rational models: Alternative algorithms for category learning. *Psychological Review*, *117*(4), 1144–1167.

Scriven, M. (1991). The methodology of evaluation. In A. A. Bellack & H. M. Kliebard (Eds.), *Curriculum and evaluation* (pp. 334–371). Berkeley, CA: McCutchan.

Shakeri, S. (2003). A mathematical modeling framework for scheduling and managing multiple concurrent tasks. Ph.D. thesis, Oregon State University, Corvallis, OR.

Shakeri, S., & Funk, K. (2007). A comparison of human and near-optimal task management behavior. *Human Factors*, *49*(3), 400–416.

Simon, H. A. (1955). A behavioral model of rational choice. *The 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.

Simon, H. A. (1990). Invariants of human behavior. *Annual Reviews in Psychology*, *41*(1), 1–20.

Simon, H. A. (1991). Cognitive architectures and rational analysis: Comment. In K. VanLehn (Ed.), *Architectures for intelligence: The 22nd Carnegie Mellon Symposium on Cognition* (pp. 25–39). Lawrence Erlbaum, Hillsdale, NJ.

Simon, H. A. (1996). *The Sciences of the artificial* (3rd ed.). Cambridge, MA: The MIT Press.

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. doi:10.1037/a0030850.

Stewart, N., Chater, N., & Brown, G. D. A. (2006). Decision by sampling. *Cognitive Psychology*, *53*(1), 1–26.

Thaler, R. H. (1994). *Quasi rational economics*. New York, NY: Russell Sage Foundation.
Todd, P. M., & Gigerenzer, G. (2001). Shepard's mirrors or Simon's scissors? *Behavioral and Brain Sciences*, *24*(04), 704–705.
Todd, P. M., Gigerenzer, G., & the ABC Research Group (2012). *Ecological rationality: Intelligence in the world*. New York, NY: Oxford University Press.
Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131.
Vaughan, W, Jr. (1981). Melioration, matching, and maximization. *Journal of the Experimental Analysis of Behavior*, *36*(2), 141–149.
Venturino, M. (1997). Interference and information organization in keeping track of continually changing information. *Human Factors*, *39*(4), 532–539.
Wilson, J. Q., & Herrnstein, R. J. (1985). *Crime and human nature*. New York, NY: Simon & Schuster.
Yechiam, E., Erev, I., Yehene, V., & Gopher, D. (2003). Melioration and the transition from touch-typing training to everyday use. *Human Factors*, *45*(4), 671–684.