children (Yu et al. 2019), as well as the quality of explanations parents provide in response to their children's questions (Kurkul & Corriveau 2018), systematically vary with several key factors of home life. A child whose parents are less likely to ask questions or provide causal explanations may thus acquire a very different-looking cost function for (e.g.,) the heuristic of reaching out to others for information than a child whose parents are more likely to engage in these kinds of behaviors. Indeed, this notion is consistent with recent computational work which suggests that learners may bring expectations about the teaching style of their informant to bear in future learning (Bass et al. 2018).

Although development provides special opportunities to employ resource-rational analysis by leveraging variability in the population, challenges remain. First, the *goals* of a developing system may radically vary from those in adulthood. For example, the goals of an adult semantic memory system might be defined by compression and storage for optimal later accessibility (e.g., Anderson 1989); however, hypothetically, a developing memory system's goal might be to expand and re-encode for representational restructuring. Because there is significantly less work that has focused on defining goals of the developing mind, resource-rational models will be underconstrained.

Second, variability within a developing child presents a challenge as algorithmic utilities are learned. According to the rational-resource analysis, the max ordered value of a heuristic depends on utilities that will be derived from representation, cognitive constraints, experiences, and rule-discovery. But these are constantly shifting in development, so how might a learner develop a preference for a particular heuristic? Consider a learner whose working memory limitations lead to favoring a "local search" heuristic. Although the learner's working memory capacity may grow over time, once a particular heuristic has been learned and habitually adopted, it is not clear when or why the system would re-evaluate and discover a more optimal "global" search heuristic employing newly developed resources. Such considerations suggest that a broader, dynamic framework of resource-rational analysis will need to be developed.

Overall, we think the resource-rational approach presented by Leider and Griffiths will be an important computational toolkit for cognitive psychology. Although there are challenges, we suggest that the variability found in cognitive development in particular will be a critical playground for modelers employing this technique.

Resource-rational analysis versus resource-rational humans

Dobromir Rahnev

School of Psychology, Georgia Institute of Technology, Atlanta, GA 30332.
rahnev@psych.gatech.edu
www.rahnevlab.gatech.edu

doi:10.1017/S0140525X19001699, e19

Abstract

Lieder and Griffiths advocate for resource-rational analysis as a methodological device employed by the experimenter. However, at times this methodological device appears to morph into the substantive claim that humans are actually resource-rational. Such morphing is problematic; the methodological approach used by the experimenter and claims about the nature of human behavior ought to be kept completely separate.

A healthy adult asked to run 60 m will likely sprint; a healthy adult asked to run 1,000 m will likely jog. In fact, there is hardly anyone on Earth who would even attempt to sprint for 1,000 m. This simple observation demonstrates what we intuitively already know: that human behavior is typically adapted to our own limitations. Therefore, a deep understanding of behavior necessitates that various sources of limitations are rigorously identified and precisely quantified. I applaud Lieder and Griffiths (L&G) for advocating for this practice.

L&G propose "resource-rational analysis," which is a methodological device that an experimenter uses to discover something about human behavior. However, the target article appears to sometimes conflate this methodological device with the substantive claim that humans are actually resource-rational. To be fair, L&G stop short of claiming that people are actually resource-rational. They even offer that "we should not expect people's heuristics to be perfectly resource-rational" (sect. 3, para. 6). Nevertheless, other parts of the target article give a sense that L&G really do think that people are (mostly) resource-rational. For example, they consider seriously the "assumption that the brain is approximately bounded-optimal" (sect. 5.3.2., para. 2), claim that resource-rational analysis "has already shed new light on the debate about human rationality" (abstract), and even state that "people's decision-mechanisms appear to be surprisingly resource-rational" (sect. 6, para. 3). These statements leave the realm of methodological devices and venture into the land of substantive claims about human rationality.

The problem is that, as currently constructed, resource-rational analysis does not and could not provide evidence for the rationality of human behavior. There are at least three reasons for this.

First, resource-rational analysis in overly flexible as a tool for establishing the nature of human behavior. As Box 2 demonstrates, a researcher who follows the methodology prescribed by L&G should test a number of different constraints and computational architectures until some combination of them provides a good fit to the data. To L&G's credit, they do advice that the experimenter stops trying out new combinations after "reasonable attempts have been made to model the constraints" (Box 2). Nevertheless, for most experimental tasks, it is not too difficult to find a set of assumptions that makes behavior to appear close to rational. This does not, however, imply that the underlying behavior is rational because the experimenter may have unwittingly postulated computational architectures or resource limitations that do not exist, or, more likely, exist but are mischaracterized. For example, a tendency to underuse explicitly stated probabilities (Rahnev and Denison 2018a) can be cast as optimal decision making by an organism that misrepresents probabilities (Zhang and Maloney 2012). However, this explanation could be given regardless of whether the organism actually adopts optimal decision making based on skewed representations of probability or adopts a suboptimal decision strategy on internal representations of probability that are less skewed. Therefore, substantive claims about human rationality require models that are prespecified and have no free parameters (e.g., the misrepresentation of probabilities should be predetermined for each subject). Very few papers, however, fit such zero-parameter models to the data.

Second, the types of tasks that we study in the laboratory tend to be the most constrained and simple tasks that an organism could ever face. Yet, even for such tasks, suboptimality is the norm (Rahnev and Denison 2018a). Regardless of how close to rationality humans get in such tasks, it does not follow that behavior would be similarly rational in the infinitely more complex real world. As L&G admit themselves, it is "challenging to [apply resource-rational analysis] to decision-making in the real world where the sets of options and possible outcomes are much larger and often unknown" (sect. 6, para. 7).

Third, the computations required to establish the truly rational strategy are intractable and will always remain so. Indeed, as demonstrated by Equation 4 in the target article, specifying what is actually rational requires quantitatively describing all environments that one has ever experienced (including environments that have been experienced by one's ancestors and have influenced brain development over evolutionary scales), which is clearly infeasible in practice. Therefore, in the strictest sense of Equation 4, we will never be able to test whether any behavior is truly rational or not.

If there is little hope that we could ever establish whether human behavior is really rational, does that mean that resource-rational analysis is also futile? Not at all. As the example of running a shorter versus longer distance demonstrates, we are profoundly constrained by our limitations, and our behavior is often roughly adapted to these limitations. Therefore, resource-rational analysis offers at least two large benefits (in addition to what was highlighted by L&G). First, resource-rational analysis can be used to *approximate* human behavior under the assumption that evolution has adapted our behavior to the particular task used by the experimenter. Clearly, for a non-resource-rational human, the approximation may be crude and sometimes very imprecise, but at the very least could be used as a starting point. Second, behavior that is systematically deviating from resource-rationality may indicate the existence of a new, previously undiscovered limitation or cognitive architecture. As highlighted above, postulating limitations just for the sake of fitting data is a dangerous undertaking, and thus any proposal for a new limitation should be tested with independent data and, ideally, under new conditions.

Regardless of one's preferred view of human nature and the best methods to reveal that human nature, it is critical that substantive claims about behavior and methodological approaches about studying said behavior are kept separate from each other (Rahnev and Denison 2018b). The person who jogs for 1,000 m is unlikely to do so at the optimal pace. That is, she is unlikely to be fully resource-rational. However, we will certainly understand her behavior better if we put in the effort to quantify the exact rate at which her muscles tire. Resource-rational analysis can be useful even if we are trying to characterize non-resource-rational humans.

Resource-rationality and dynamic coupling of brains and social environments

Don Ross[a,b,c]

[a]School of Sociology, Philosophy, Criminology, Government, and Politics, University College Cork, Cork T12 AW89, Ireland; [b]School of Economics, University of Cape Town, Rondebosch 7701, South Africa and [c]Center for Economic Analysis of Risk, J. Mack Robinson College of Business, Georgia State University, Atlanta, GA 30303.
don.ross931@gmail.com
http://uct.academia.edu/DonRoss

doi:10.1017/S0140525X1900150X, e20

Abstract

Leider and Griffiths clarify the basis for unification between mechanism-driven and solution-driven disciplines and methodologies in cognitive science. But, two outstanding issues arise for their model of resource-rationality: human brains co-process information with their environments, rather than merely adapt to them; and this is expressed in methodological differences between disciplines that complicate Leider and Griffiths' proposed structural unification.

Leider and Griffiths' (L&G) project, to offer an explicit framework for relativizing assessments of rationality simultaneously to cognitive processing constraints and environmental affordances, represents important progress. It significantly clarifies the basis for unification between mechanism-driven and solution-driven disciplines and methodologies, as they say. But, as the framework is extended and refined, two outstanding issues merit consideration: (1) human brains do not merely adapt to their environments, but co-process information with their environments, particularly with its social aspects; and (2) L&G's idealization of disciplines as standing in a hierarchy of abstraction from mechanism details is a somewhat misleading simplification of methodological reality.

L&G's core Equation 4 takes the environment (E) as a fixed constraint on optimal heuristic selection. This is reasonable in light of the long time-scale for learning that their discussion indicates that they have in mind, reflected in their comment that evolution and cognitive development "solve the constrained optimization problem defined in Equation 3" (sect. 3, para. 5). The framework obviously allows for environmental variation, across time or space, to be modeled and analyzed using comparative statics. Furthermore, their inclusion of the information term I on the left-hand side of Equation 4 recognizes that learning encoded in the genome is refined by learning in the phenome. However, the model seems to presuppose that cognitive processing is all done in the brain, because there is no interaction term involving all of h, E, and B (heuristics, environment, and brain).

This may be a reasonable idealization where most cognitive systems are concerned. But, it might be seriously misleading in the case of humans equipped with writing, art, and mathematics, who have populated their environments with technologies that actively process information in conjunction with inboard cognition. Obvious examples include external computing devices, but these are not the main source of potential deep complication for L&G's model. Though, the relationship between a person and a machine she uses may be dynamically interactive, in non-exotic cases the extent of such dynamical coupling is both limited and specifiable; and, as noted above, this is all that is required for analyzing variation by means of comparative statics. The more serious challenge arises from the abstract technology of social institutions. Ecologically, humans are arguably most strongly distinguished from other highly intelligent animals by their use of shared information-processing routines that are encoded in rules, norms, and institutionalized procedures, which