Samuli Poyhonen (samuli.poyhonen@helsinki.fi), University of Helsinki
Contributed paper, PSA2016
Manuscript date 1.3.2016

What, when and how do rational analysis models explain?

(Word count: 4995)

Abstract

Probabilistic modeling is a highly influential method of theorizing in cognitive science. Rational analysis is an account of how probabilistic modeling can be used to construct non-mechanistic but self-standing explanatory models of the mind. In this article, I disentangle and assess several possible explanatory contributions which could be attributed to rational analysis. Although existing models suffer from evidential problems that question their explanatory power, I argue that rational analysis modeling can complement mechanistic theorizing by providing models of environmental affordances.

1. Introduction

During the past two decades, probabilistic modeling has become one of the most visible strands of cognitive modeling alongside connectionism, rule-based approaches and dynamical systems modeling. Curiously, against the general trend in the cognitive sciences where theorizing is increasingly anchored in neuroscience findings, probabilistic modeling of higher cognition has been a characteristically top-down endeavor. Without making any substantial commitments about the underlying cognitive mechanisms, probabilistic modeling has been applied to complex aspects of human cognition, which have often been thought of as being beyond the reach of mechanistic research methods. Models of human memory, categorization, causal learning, concept learning, and conditional inference, to mention a few applications, often show an impressive fit with empirical

data, and the novel analyses of cognitive capacities provided by the models appear to have shed new light on the nature of the explananda under study.

However, how does that shedding light actually occur – how do such computational probabilistic models explain? Although probabilistic modeling, in principle, does not rely on any particular method of explanation, modelers often refer to the idea of rational analysis as the account of how and why their models help us understand the mind (Anderson 1990; Oaksford & Chater 2007). The striking claim made by rational analysis (RA) modelers is that by understanding higher cognitive capacities as forms of inductive inference, we can predict behavior, and understand a lot about human cognition without making any assumptions about the underlying representations and processes. This agnosticism about neural and cognitive mechanisms is justified by making reference to the rationality of human behavior: We know that human agents tend to be generally well-adapted to their environment, and hence a careful analysis of the cognitive task encountered by the mind, coupled with an assumption of the optimality of human behavior, results in a putatively powerful methodology of prediction and explanation.

However, there is a large consensus in the philosophy of science that explanations also in the cognitive sciences should track causal mechanisms, and the way RA purports to sidestep the evidential and explanatory problems arising from the causal complexity of cognition has given rise to a strongly polarized debate (see, e.g., peer commentary in Jones & Love 2011). On the one hand, the way that the new mathematical methods in probabilistic modeling can combine structure and learning in human thought has lead to an exciting new paradigm for theorizing about the mind. On the other hand, the proponents of non-causal explanation need to show when and how it is that non-causal models explain rather than redescribe or merely formally unify various phenomena (cf.

Colombo & Hartmann 2015). Otherwise, rational analysis could simply be seen as the last breath of the autonomist dream of studying the mind independently from the brain.

In this paper, I assess the explanatory status of RA models by disentangling various explanatory contributions which have been attributed to them. By relying on the contrastive-counterfactual theory of explanation, I distinguish between three possible explanatory contributions such models could make: Uncovering (a) constitutive dependencies between parts and wholes, (b) environment-behavior dependencies, and (c) environment–optimal behavior dependencies. I treat the third alternative as the most promising source of new understanding provided by RA models. I argue that (c) should be interpreted as being explanatory not of human behavior as such, but of environmental affordances. Well conducted modeling of environmental affordances can complement mechanistic theorizing by providing means for understanding the possible space of behavior of agents.

2. Probabilistic cognitive modeling and rational analysis

2.1 Procedure of rational analysis

The idea of rational analysis modeling dates back to John Anderson's work on human memory and categorization in *The Adaptive Character of Thought* (1990). Having already worked on his ACT* cognitive architecture, the new methodology put forward in the book reflected Anderson's increasing worries that the research methods of the time could not really uncover cognitive mechanisms. Lacking a clear picture of what it is that cognitive mechanisms do (i.e. what the psychological explananda are), the available evidence of neural and algebraic level structures was insufficient to uncover the mechanistic architecture of the human mind (Anderson 1990, pp.23–26). Compared to bottom-up research strategies, rational analysis begins from the other end:

[…] We can understand a lot about human cognition without considering in detail what is inside the human head. Rather, we can look in detail at what is outside the human head and try to determine what would be optimal behavior given the structure of the environment and the goals of the human. (Anderson 1990, p.3)

According to Anderson, careful mathematical modeling of the environment/task structure combined with an assumption about the optimality of human behavior leads to a new self-standing research strategy for understanding the mind: *"As this book is evidence, a rational analysis can stand on its own without any architectural theory"* (ibid.). By providing a precise model of what the mind does as a well-adapted system, rational analysis can constrain the search space for cognitive mechanisms, and put the scientific study of the mind on a firm foundation.

This view of the role of computational modeling immediately brings to mind Marr's (1982) account of multi-level theorizing in the mind sciences. However, whereas Marr provides no systematic model for building computational-level theories, RA modeling has predominantly proceeded according to the six-step modeling cycle proposed by Anderson (1990. p.29):

1. Specify precisely the goals of the cognitive system
2. Develop a formal model of the environment to which the system is adapted
3. Make minimal assumptions about computational limitations
4. Derive the optimal behavior function, given items 1 through 3
5. Examine the empirical evidence to see whether the predictions of the behavior function are confirmed
6. Repeat, iteratively refining the theory

These steps embody an account of how a large part of probabilistic cognitive modeling is done. However, two further assumptions should be made explicit. First, the derivation of optimal behavior in steps 2-4 typically employs *probability calculus* (not logic) as the normative baseline theory of rational behavior. Secondly, the connection between model predictions (step 4) and observed behavior of humans (step 5) is mediated by an assumption about the *optimality of the observed behavior* (see quoted passage above).

Below I illustrate this process with an example. However, a comment on the status of the approach in cognitive science is in place: Not all probabilistic modelers endorse the rational analysis framework (cf. Danks 2015; Sakamoto et al. 2008; Brighton & Gigerenzer 2008). Focusing on RA is useful for two reasons, however. Rational analysis is undeniably influential, and its core commitments have been endorsed a large group of well-known modelers (e.g., Anderson 1990; Oaksford & Chater 1994, 2007; Griffiths & Tenenbaum 2009). A further advantage of focusing on RA has to do with the fact that often the theoretical commitments of mathematical modelers are hard to pin down. In some cases, this is surely due to the modelers themselves not being clear of where their commitments (about explanatoriness, optimality, etc.) lie. Rational analysis provides a clear account of the conceptual foundations of probabilistic cognitive modeling, and therefore the following discussion is potentially helpful for challenging the methodological quietism among probabilistic modelers.

2.2 Oaksford and Chater on the Wason selection task

To illustrate the rational analysis process, I now briefly introduce Mike Oaksford and Nick Chater's (1994, 2007) analysis of the Wason selection task. Being a relatively simple model, it is a good device for illustrating the conceptual basis of RA modeling.

Wason selection task is one of the most famous laboratory experiments discussed in the literature on human rationality. In the original form of the task, subjects are given four cards, each of which has a letter on one side and a number on the other. The subjects' task is to determine whether the rule *"If there is a vowel on one side of the card (p), then there is an even number on the other side (q)"* holds. More precisely, subjects are asked to select all those cards, but only those cards, which would have to be turned over in order to discover whether the rule is true for the combination of cards they were given. The famous finding from the task and its several replications is that only a small minority of the subjects (less than 10%) select the correct cards (vowel, odd number) corresponding to the falsifying instance. Judged in the light of logic, most subjects fail to perform in a rational way.

Oaksford and Chater (O&C) challenge the irrationality claim by arguing that logic-based theories of inference and rationality misrepresent people's behavior in the task. O&C's own *information-gain model* of the situation argues that the apparently irrational behavior can be understood as the optimal way of decreasing uncertainty regarding the hypotheses studied. The gist of O&C's reinterpretation of the selection task is that instead of engaging in deductive reasoning, subjects interpret the task as inductive one. They do not try to falsify the rule, but instead they try to determine which of two hypotheses holds:

(a) Independence hypothesis **H_i:** $P(q \mid h) = P(q)$ or

(b) Dependence hypothesis **H_d:** $P(q \mid p)$ is high, higher than $P(q)$.

Being initially equally uncertain about both hypotheses, subjects aim to reduce this uncertainty as much as possible by turning as few cards as possible.

The rational analysis proposed by O&C relies on three basic starting points:

(1) Higher cognition can be modeled as probabilistic (Bayesian) computation

(2) The likelihoods and prior probabilities required by the model can be acquired from the analysis of the environment structure

(3) Behavior of human agents constitutes an optimal response to the task.

The Bayesian model of the situation is constructed roughly as follows.[1] To formalize the idea of uncertainty reduction, O&C adopt the optimal data selection paradigm, and interpret uncertainty reduction as optimization of *expected information gain.* Expected information gain $E[I_g]$ from turning over a card is defined as $E[I(H_i|D) - I(H_i)]$.[2] The Shannon information terms $I(H)$, in turn, are a function of the probabilities of the hypotheses before and after observing data, $P(H_i)$ and $P(H_i|D)$. These required posterior probabilities can be calculated from the likelihoods $P(D|H)$ and the priors by applying the Bayes rule. As the initial priors were set to be equal (.5), the rest of the crucial model specification is built into the likelihood functions, which describe the nature of the four-card task. Oaksford and Chater (1994, Table 1) show in detail how the required likelihoods can be read off the contingency tables describing the two hypotheses.

From these derivations, it follows that the crucial parameter values determining the optimality of behavior are the base rates of p and q. These probabilities describe how often positive instances of

---

[1] For mathematical details, see Oaksford & Chater 1994, 2007.

[2] Uncertainty (Shannon information) $I(H_i)$ given n mutually exclusive and exhaustive hypotheses (H_i), is $-\sum_{i=1}^{n} P(H_i) \log_2 P(H_i)$.

the antecedent and consequent of the rule appear in the environment. The expected information gain from turning the four cards depends on P(p) and P(q) in the following way:

- P(q) is small → P card is informative
- P(p) is large → Not-q card is informative
- P(p) and P(q) are small → Q card is informative
- Not-p card is not informative

How should these base rates, then, be determined? Instead of attempting to somehow measure the base rates of vowels and consonants in a relevant environment, O&C cite various intuitively plausible justifications for their *rarity assumption*. Relying on the observation that categories in language cut the world quite finely, the rarity assumption states that, generally, P(p) and P(q) are low in most situations.[3] Under rarity, O&C conclude, the q card is more informative than the not-q card. Hence, the model concludes that highest expected information gain is achieved by turning p and q cards, exactly as a majority of the participants do. Actually, with the parameter values chosen by O&C, there's a very good fit between meta-analysis results about people's behavior in the standard form of the selection task, and the predictions of the model. Hence, by changing the normative model of rational behavior, O&C were able to explain away irrationality, and to show that experimental subjects' behavior is actually very close to optimal.

The model has received critical attention in the literature (cf. Oaksford & Chater 2009), but it serves our current purposes well. The model specification and the modeling assumptions are conceptually

[3] See Oaksford & Chater 1994, 2007, and 2009 for alternative justifications of rarity.

on a par with those in more complex Bayesian models. The complexity in such models often pertains to the structure and generation of hypothesis spaces, and the models often rely on computational tools (such as MCMC approximation methods) to make the calculations tractable. However, these mathematical complexities have no influence on the fundamental conceptual structure of the model. What is common to all such models is that the none of the components (hypothesis space, likelihood function, and priors) are interpreted in a psychologically realistic way as mental representations (Jones & Love 2011). Instead, they stand directly for properties of the environment. Furthermore, data about human behavior is not fed into the model specification to empirically calibrate the model. Instead, it is only used to test model predictions. Hence, in this sense, the rational analysis of the selection task is an illuminating example of the theoretical and conceptual assumptions of computational probabilistic modeling.

3. What rational analysis models fail to explain

A shared starting point for many accounts of scientific explanation has been to distinguish explanation from other epistemic activities (e.g., description and prediction) by pointing out that explanations offer information of a specific kind. Explanations show *how* or *why* something happened or obtains. According to a now widely accepted approach, the knowledge that allows one to answer such questions concerns change-relating counterfactual dependencies between the relata in the explanation.

Stated generally, according to this contrastive-counterfactual theory of explanation, explanatory information has the following form (Woodward 2013; Ylikoski & Kuorikoski 2010):

{CC} *x [x'] because of y [y']* (variable X takes the value x instead of x' because Y has the value y instead of y')

In this account, being able to explain can be captured by being able to correctly answer what-if-things-were-different questions, i.e. questions of how changes in explanans variables lead to changes in the explanandum variable. In addition to being a sufficiently general account of explanation, the contrastive-counterfactual theory suits the purposes of this article well, because it does not necessarily tie the notion of explanation to that of causation. That is, although the 'because' in {CC} is typically understood as referring to causal dependency, the account does not rule out the possibility of non-causal explanation (Woodward 2013; Pincock 2015; Rice 2015): If there are ways of defining the notion of invariant dependency in non-causal situations (e.g. for mathematical dependencies), the contrastive-counterfactual theory could be applied to non-causal explanations as well. Hence, the theory of explanation casts the net wide enough to give RA models a fair chance of being explanatory.

A further advantage of treating explanations as answers to questions is that it allows us to make more precise the possible explanatory claims made by RA modelers. I suggest that there are at least three different kinds of objective dependencies that RA models could be said to track: (1) constitutive dependencies between parts and wholes, (2) environment-behavior dependencies, and (3) environment–optimal-behavior dependencies. In the rest of this section, I argue that in most cases of RA modeling, there are good reasons to conclude that the models do not have genuine explanatory import with respect to the two first kinds of dependencies.

3.1 Constitutive what-ifs

The notion of mechanism has acquired a central position in the philosophical debates concerning explanation in the life sciences. A clear expression of the mechanistic viewpoint has recently been given in the *model-to-mechanisms mapping (3M) requirement* by Kaplan and Craver (2011). According to the requirement, dynamical and mathematical models in systems- and cognitive

neuroscience explain a phenomenon only if there is a mapping between elements in the model and elements in the mechanism for the phenomenon. As the example discussed above suggests, rational analysis models provide no such mapping. They are agnostic about algorithmic and implementation level details, and intentionally so. Does this mean they cannot be explanatory?

First, as Kaplan and Craver themselves admit, their argument ultimately relies on shared norms about explanatoriness in the neuroscience community, and their account of explanation as construction of multi-level mechanisms reflects these norms. However, if such norms do not hold among probabilistic cognitive modelers, it is not obvious why they should abide by the 3M requirement.

Instead, if we understand explanation according to the contrastive-counterfactual theory, Kaplan and Craver's argument seems less disastrous: RA models obviously do not provide information about constitutive and causal dependencies in multi-level mechanisms, but according to the account, this does not rule out the possibility of RA models tracking some other kinds of objective dependencies, e.g. those holding between relata described in computational-level terms. Furthermore, a proponent of RA need not (and should not) claim that adding mechanistic detail never improves a computational explanation. To defend explanatoriness of RA models, a far weaker claim suffices, one stating that there can be explanatory contributions which do not rely on information from uncovering causal mechanisms.

3.2 Environment–behavior what-ifs

A second kind of explanatory question answered by an RA model could be "how would the behavior of the cognizer change when the cognitive task changes in some particular way?" That is, a RA model could uncover objective dependencies between properties of the environment and the

behavior of cognizers. For example, O&C's model can be used to derive predictions of what the behavior of the subjects in the Wason tasks would be, were P(p) and P(q) to take a range of values.

It is here that the optimality assumption of RA becomes crucial. To predict how human behavior would change in response to changes in the task, without knowing anything about the algorithms and processes which produce behavior, RA relies on the assumption that humans are well-adapted to their environments: If we assume that human behavior is optimal (or approximates optimal behavior) across a large variety of environments, the predictions derived from the RA model (step 4 of the analysis procedure) should in fact apply to that behavior.

Given that human (ir)rationality has been the topic of a longstanding debate in philosophy and psychology, it is not surprising that the optimality assumption has drawn a lot of criticism (cf. Jones & Love 2011). Although proponents of RA are correct in arguing that some degree of rationality of target behavior is required for us to even perceive it as intentional action, the modest levels of rationality needed hardly license the strong optimality assumptions in RA models. Neither do evolutionary arguments provide support for strong optimality claims: Although natural selection is a source of design and adaptedness, evolution is not guaranteed to produce globally optimal solutions – merely a local comparative advantage is sufficient for evolutionary solutions to survive.

Being aware of these problems, proponents of RA have avoided appealing to evolutionary defenses of the optimality assumption. Instead, they justify optimality by relying on an analogy to behavioral ecology and economics, where similar assumptions are commonly made (Chater et al. 2003). I believe, however, that the analogy breaks down due to a crucial dissimilarity between these fields: Both in biology and economics, rationality claims typically concern aggregate behavior, not that of

individual agents. Due to the disanalogy, I do not see how appealing to economics or biology could be a viable way to justify optimality assumptions in RA modeling.

These problems with general defenses of the optimality assumption suggest that perhaps optimality should be examined more locally. What kind of evidence should be obtained to justify the optimality claim in the case of a particular cognitive task? It seems that to support an objective dependency between environment and behavior, we should gather data about human behavior in a task *across a range of parameter values* describing various different environmental states. If human behavior fits the predictions made by the model across a range of conditions, that would appear to be rather strong evidence of optimality.[4]

Existing RA models rarely employ such cross-environmental data. First of all, many models not rely on any actual measurements of environment parameters (cf. Jones & Love 2011). Instead, they use plausible-sounding assumptions or analogies. For example, in O&C's selection task model, the base rates for p and q originated in such analogical reasoning. Similarly, Anderson's (1990, ch. 2) early model of memory relied on data about library borrowings to model usage of memory structures, and Griffiths et al. (2007) use Google PageRank to predict fluency of recall. Models devoid of good quality empirical data should be considered as toy models (at best), incapable of uncovering actual properties of cognitive environments.

---

[4] Note, however, that such empirical evidence for optimality would make the theory-based optimality assumption unnecessary.

Furthermore, as Marcus and Davis (2013, Table 1) observe, Bayesian modelers have been selective in the results that they report from experimental tasks. They only report ones where human behavior follows the model and ignore cases where its not optimal. Although some of the most recent models show some improvement in these respects, generally in RA models there is little evidence that could support knowledge of the needed invariant environment-behavior counterfactuals.

4. Rational analysis and the logic of the situation

Finally, let us think about the epistemic value of a RA model if we drop the optimality assumption. Assume that we have a rational analysis model with (i) well-specified task structure, (ii) parameter values based on empirical measurement of the environment, and (iii) an account of computational costs and limitations. What such a model could do is to link combinations of parameter values to best possible behavioral choices in those situations. Is this not a kind of objective what-if dependency? However, consider what the relata of such a dependency are. The model tells what the optimal behavior would be, given a particular combination of environmental conditions and computational limitations. Such counterfactuals do not say anything about actual human behavior. Instead, they increase our understanding of the environmental affordance, or, the logic of the situation (Popper 1963).

What mathematical models of affordances – the opportunities the environment offers for the agent – can help us understand is the possible space of behavior for cognitive agents. They show what a hypothetical rational agent would do in different situations. For what purposes could such information be useful? First, were we to design artificial cognitive systems with a particular cognitive task in mind, these systems should approximate the optimal behavior specified by the model. For example, in the selection task, *if* we are interested in reducing our uncertainty, O&C's

model tells us something non-trivial: It reveals the best choices of cards under different values of base rates for p and q.

Secondly, as in economics, rational models can act as normative baselines to which human behavior can be compared. As Sloman & Fehrbach (2008) argue, often it is just as interesting to find out that behavior does not conform to the norm than when it does. Finding out where and how systems malfunction is an efficient way to learn about them.

However, in neither of these uses are RA models employed to directly explain human behavior. Instead, they function as inferential aids which help to map the possible space of action for agents when faced with a particular task. Herein lies perhaps the hardest evidential problem faced by rational analysis. How do we know what the mind really does in some situation, i.e. where do the functional hypotheses in step 1 of RA come from? For example, how would O&C defend their probabilistic construal of the selection task against an adamant falsificationist? Available empirical evidence can hardly decide the issue: Where O&C see optimal behavior, the falsificationist sees well-known inferential blunders.[5] Marcus and Davis (2013) argue that similar problems of model selection plague several other RA models as well.

The difficulty seems to come down to the fact that the cognitive tasks and the affordances available for an organism depend on its "life space" – not the physically objective world in its totality, but reality filtered through the organism's needs, drives and perceptual apparatus (Simon 1956).

---

[5] What makes O&C's model selection seem even more ad hoc is that they do not explain different versions of the selection task (e.g., the deontic selection task) by using the same model, but instead they introduce modified versions for each of the variations.

Therefore, there is no reason to think that a mathematician's intuitions are a reliable guide to what the cognitive tasks of human agents are. Ad-hocness in model selection, in turn, raises serious worries about the *relevance of RA modeling*: Constructing detailed mathematical models of potential affordances is of little interest unless they can be shown to be ones humans actually track.

This leads me to my conciliatory conclusion. As suggested both by the connectionist rivals of RA and proponents of multi-level mechanistic explanation in philosophy (McClelland et al. 2010; Bechtel & Richardson 2010), functional hypotheses in cognitive science must be formulated in an iterative process between bottom-up and top-down research strategies. On the one hand, knowledge about perceptual and computational constraints of organisms mostly originates in bottom-up research on the mind-brain, and this knowledge should be allowed to constrain RA models. In this sense, Anderson's and O&C's claims about the self-standing explanatory role of RA are not vindicated by my analysis. However, the discussion on mechanistic explanation has been downward-looking in spirit, and modeling the environment within which cognitive mechanisms function has not received enough attention. Here RA models can complement mechanistic theories of cognition by providing precise mathematical models of the task and the environment. For example, as Chater et al. (2003) point out, a correctly formulated rational analysis can show why it is that some simple approximating heuristic is successful in solving a computationally complex task.

4. Conclusions

I have argued that given a sufficiently broad account of scientific explanation, there are several possible ways in which probabilistic modeling could increase our understanding of the mind. However, the strictly-computational methodology embodied in the six-step formula of rational

analysis has led to theorizing which often fails to reliably uncover genuine explanatory dependencies. The shortcomings of RA are evidential in nature: The nature of the data, and the way it is used in model construction allows too easy curve fitting, and it is insufficient for reliable counterfactual inference.

My new proposal about the epistemic role of RA models without the problematic optimality assumption is that they can be understood as models of environmental affordances. Interpreted in this way, RA models do not actually provide information about the mind works, or hypotheses about cognitive functions (Zednik & Jäkel 2014). Instead, they map the possible cognitive space of action for an organism. The explanatory contribution of such information is best worked out as constituting a part of a non-reductionist mechanistic research programme.

References

Anderson, John. 1990. *The Adaptive Character of Thought*. Hillsdale: Lawrence Erlbaum Associates.

Bechtel, William. and Richardson, Robert. 2010. *Discovering Complexity.* The MIT Press.

Brighton, Henry. & Gigerenzer, Gerd. 2008. Bayesian brains and cognitive mechanisms: harmony or dissonance? In Chater & Oaksford (eds.) *The Probabilistic Mind*. Oxford University Press.

Chater, Nick, et al. 2003. Fast, frugal, and rational: How rational norms explain behavior. *Organizational Behavior and Human Decision Processes,* 63–86.

Chater, Nick. & Oaksford, Mike. (eds.) (2007). *The Probabilistic Mind.* Oxford: Oxford University Press.

Colombo, Matteo. & Hartmann, Stephan. (2015). Bayesian cognitive science, unification, and explanation. *British Journal for the Philosophy of Science.*

Danks, David. 2015. *Unifying the Mind*. MIT Press.

Griffiths, Thomas., Steyvers, Mark., & Firl, Alana. 2007. Google and the mind. *Psychological Science*, 1069–1076.

Griffiths, Thomas. & Tenenbaum, Joshua. 2009. Theory-based causal induction. *Psychological Review*, 661–716

Jones, Matt. & Love, Bradley. 2011. Bayesian Fundamentalism or Enlightenment? *Behavioral and Brain Sciences*, 34, 169-231.

Kaplan, David. & Craver, Carl. 2011. The explanatory force of dynamical and mathematical models in neuroscience: A mechanistic perspective. *Philosophy of Science*, 78, 601-627.

Marcus, Gary, & Davis, Ernest. 2013. How robust are probabilistic models of higher-level cognition? *Psychological Science*, 24, 2351–2360.

Marr, David. 1982/2010 *Vision*. W.H. Freeman/MIT Press.

Oaksford, Mike, & Chater, Nick. 1994. A rational analysis of the selection task as optimal data selection. *Psychological Review*, 101, 608-631,

— 2007. *Bayesian Rationality: The probabilistic approach to human reasoning*. Oxford University Press.

— 2009. Precis of Bayesian Rationality. *Behavioral and Brain Sciences*, 69–120.

Pincock, Christopher. 2015. Abstract explanations in science. *British Journal for the Philosophy of Science* 66, 857-882.

Popper, Karl. 1963. Models, instruments, and truth. Manuscript. Karl Popper Collection at the Hoover Institution Archives at Stanford University.

Rice, Collin. 2015. Moving beyond causes: Optimality models and scientific explanation. *Noûs* 49, 589-615.

Sakamoto, Yasuaki., Jones, Matt. & Love, Bradley. 2008. Putting the psychology back into psychological models**.** *Memory & Cognition*, 36, 1057–1065.

Simon, Herbert. 1956. Rational choice and the structure of the environment. *Psychological Review*, 129–138.

Sloman, Steven, & Fehrbach, Philip. 2008 The value of rational analysis: as assessment of causal reasoning and learning. In *The Probabilistic Mind.*

Woodward, James. 2013. Mechanistic explanation: Its scope and limits. *Proceedings of the Aristotelian Society Supplementary Volume,* lxxxvii: 39–65.

Ylikoski, Petri., & Kuorikoski, Jaakko. 2010. Dissecting explanatory power. *Philosophical Studies*, 148, 201–219.

Zednik, Carlos. & Jäkel, Frank. 2014. How does Bayesian reverse-engineering work? In P. Bello et al. (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 666-671).