How to Explain Behavior?

Gerd Gigerenzer

Harding Center for Risk Literacy, Max Planck Institute for Human Development

Abstract

Unlike behaviorism, cognitive psychology relies on mental concepts to explain behavior. Yet mental
processes are not directly observable and multiple explanations are possible, which poses a challenge
for finding a useful framework. In this article, I distinguish three new frameworks for explanations that
emerged after the cognitive revolution. The first is called *tools-to-theories*: Psychologists' new tools for
data analysis, such as computers and statistics, are turned into theories of mind. The second proposes
as-if theories: Expected utility theory and Bayesian statistics are turned into theories of mind, describ-
ing an optimal solution of a problem but not its psychological process. The third studies the *adaptive
toolbox* (formal models of heuristics) that describes mental processes in situations of uncertainty where
an optimal solution is unknown. Depending on which framework researchers choose, they will model
behavior in either situations of risk or of uncertainty, and construct models of cognitive processes or
not. The frameworks also determine what questions are asked and what kind of data are generated.
What all three frameworks have in common, however, is a clear preference for formal models rather
than explanations by general dichotomies or mere verbal concepts. The frameworks have considerable
potential to inform each other and to generate points of integration.

Keywords: Adaptive toolbox; As-if theories; Computation; Explanation; Heuristics; Tools-to-theories;
Cognitive processes; Black-box theories

1. Introduction

Children persistently ask "why?" Some pose more "why" questions than their parents can answer. Being curious about causes rather than mere associations is characteristic of human intelligence and contrasts with big data analytics, which is bound to search for correlations rather than causes. Like children, psychologists also ask "why" questions. Yet an explanation in psychology involves more than distinguishing between correlations and causes. The challenge is more fundamental: to find a theoretical language for the causes of behavior in the first place. An observed behavior can always be explained in multiple ways. For instance, behaviors such as not eating and not sleeping have been attributed to burnout, mobbing by peers, learned helplessness, inward directed anger, and neurotransmitter imbalance in the brain. The problem is to find a strong theoretical framework to narrow down the infinite number of possible explanations.

In this article, I focus on cognitive psychology and on three frameworks of explanations that emerged after the "cognitive revolution" of the 1950s and 60s. These frameworks have several striking features in common. Across the long history of psychology, they were virtually absent in the study of perception, memory, attention, and thinking. All three promote the use of formal models rather than verbal statements, and they played a role in overthrowing behaviorism. Most remarkably, they have changed the very way we think about the nature of cognition.

My goal is not to provide a taxonomy of explanations in psychology or an overview of philosophical treatises on explanation since Aristotle and Hume. The three frameworks I present are also not exhaustive but instead only characteristic for cognitive psychology. They can be classified by two dimensions: whether they deal with situations of risk or uncertainty, and whether they model cognitive processes or not. The choice of framework influences how we think of cognition. Moreover, the frameworks determine what research questions we ask and what data we generate.

I begin with the period before the cognitive revolution.

2. The black box

Imagine a black box, with sensory input entering the box from the left and behavior exiting from the right (Fig. 1). The sensory input may be visual or acoustic, and the behavior may be a physical action or a verbal statement. The most radical form of explaining observed behavior is to treat the mind as such a black box. That eliminates explanations by intentions, beliefs, or thought. Behavior on the right side of the box is explained solely by the pattern of the sensory input on the left side. In other words, the environment controls behavior. There is a time-honored justification for keeping the black box shut: observations only, without any speculations. Only that which is verifiable through observation counts as explanations, meaning that mental processes are excluded. This verdict is the fundamental position of the logical positivism of Ernst Mach, which
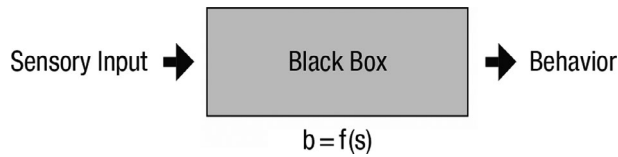
Fig. 1. The mind as a black box. Behavior (*b*) is explained by the temporal and spatial structure of the sensory input (*s*).

B. F. Skinner adopted in what he termed *radical behaviorism*. Behavior is explained by the experimental analysis of the properties of the input, resulting in laws of the form $b = f(s)$, where *b* stands for the observed behavior and *s* for the pattern of sensory input.

2.1. Example 1: Classical conditioning

Pavlov accidentally discovered classical conditioning when studying the nature of reflexes in dogs. In a reflex, an unconditioned stimulus (e.g., perceiving food) elicits an unconditioned response (salivating). When a neutral stimulus (clicks of a metronome) is paired with the food, after a number of trials, the clicks alone will cause an increase in salivation. Here, a surprising behavior (that a dog salivates when it hears a click) is explained solely on the basis of the sensory input, that is, the temporal contingency of a neutral stimulus with an unconditioned stimulus.

2.2. Example 2: Operant conditioning

Skinner showed that the speed of acquiring new behaviors and their resistance to extinction is a function of the reinforcement schedule. For instance, continuous reinforcement speeds up learning of a behavior, whereas partial reinforcement leads to behavior that is more resistant to extinction. Both the learning of new behavior and its resistance to extinction are explained by the temporal structure of the sensory input alone, such as continuous versus partial reinforcement.

Classical and operant conditioning were successful in explaining, shaping, and predicting behavior in the laboratory, and they also led to applications in animal training and human behavior therapy. The laws of conditioning provide not only explanations but also methods for controlling behavior that have inspired literary utopias and real political systems. Aldous Huxley's *Brave New World* features "neo-Pavlovian conditioning rooms" where infants from lower castes are classically conditioned to fear books and plants by associating these with the shrill noise of sirens and mild electric shocks. These children avoid books and going outdoors for the rest of their lives. In *Walden II*, Skinner proposes his own vision of a future society that uses behavioral engineering by the community as opposed to by parents. Positive reinforcement of desirable social behavior creates a peaceful world without crime, and Walden II's "credit system" rewards its members in a fair way with points. Skinner's fictional system anticipates the real *social credit system*

that China announced in 2014, which rewards "sincere" behavior and punishes insincere behavior in a drive to fight corruption and selfish behavior and build a harmonious socialist society. Similar to a Skinner box, where computerized reinforcement schedules shape the behavior of pigeons, modern digital surveillance technology allows tight monitoring and modification of the behavior of citizens. In these systems, the citizen is treated as a black box.

Behaviorism came under attack in the 1950s and 60s. For instance, John Garcia and colleagues showed that when the taste of flavored water is followed by experimentally induced nausea, rats can learn to avoid the flavored water in just one trial, but when the same taste is paired with an electric shock, rats have great difficulty learning to avoid the flavored water (Garcia & Koelling, 1966). The laws of conditioning cannot explain this difference; the rat is biologically prepared to learn specific associations, such as food and nausea, but not others. Possibly more influential than experimental research in the decline of behaviorism was Chomsky's (1959) review of Skinner's *Verbal Behavior*. His "poverty of the stimulus" argument states that learning a language— that is, verbal behavior—cannot be explained by operant conditioning alone. Instead, there must be a partly innate linguistic capacity that guides the ability to generalize and process information.

The "cognitive revolution" of the 1950s and 60s eventually overthrew behaviorism.[1] This posed the challenge to open the black box. How could a language be found for the cognitive processes that mediate the influence of the sensory input? In what follows, I distinguish three answers to this question. I will refer to these frameworks as *tools-to-theories*, *as-if theories*, and *adaptive toolbox theories*.

3. Tools-to-theories

The cognitive revolution was more than an overthrow of behaviorism and a revival of the mental—it changed the very explanation of the mental. For instance, present-day textbooks in psychology take it for granted that cognition is computation, particularly statistical computation. What else could it be? However, a glance into textbooks before the cognitive revolution shows that the concept of explaining perception, attention, memory, or thinking in terms of statistical computation is virtually absent (Gigerenzer & Murray, 2015). Explanatory concepts before the 1950s included sensory thresholds, Weber's and Fechner's laws, the laws of association, Tolman's mental maps, and Gestalt processes such as restructuring and insight. The role of probability theory was largely restricted to capturing measurement error and other forms of unsystematic error, as in Thurstone's law of comparative judgment.

Where did the new view of cognition as computation come from? It emerged after new tools for data processing—statistics and computers—were introduced into the psychological laboratories. This process of theory construction is described by the *tools-to-theories heuristic* (Gigerenzer, 1991):

1.. *Discovery*: New scientific tools, once entrenched in a scientist's daily practice, suggest new theoretical analogies and explanatory concepts.
2.. *Acceptance*: Once proposed by an individual scientist (or a group), the new metaphors and concepts are more likely to be accepted by the scientific community if their members are also users of the new tools.

In this way, new theories can be inspired directly by new tools rather than by new data. The new theories assume that cognition intuitively relies on the same tools as researchers do. As a result, cognitive processing mirrors data processing. Yet the tools-to-theories argument also has a social component: The new theories are unlikely to be accepted by the community unless its members are also familiar with the tool.

The resulting vision of the mind is shown in Fig. 2. Behavior is no longer explained by the sensory input alone, but by the input transformed by a data processing tool. Importantly, just as computers and statistical methods require a specific input to work with, the new theories also changed both the kind of input presented to experimental participants and the data produced. The sensory input itself is determined by what the tool can process.

3.1. Example 1: Signal detection theory

If the difference between two tones is smaller than 1 Hz (or 3 Hz for sine waves), people tend to perceive two tones as identical, otherwise as different. Since the 19th century, psychologists have explained this striking phenomenon (which leads to violations of transitivity) by a "differential threshold," that is, that the two stimuli need to differ by more than a "just noticeable difference" (jnd). The jnd's were considered to be the elements of the mind; the psychologist Edward Titchener counted some 44,000 of these (Gigerenzer & Murray, 2015). In the wake of the cognitive revolution, however, psychologists began to explain the same phenomenon in a fundamentally different way: Like a statistician, the mind makes an *inference* about whether the two tones are the same. Specifically, according to signal detection theory, the mind computes two sampling distributions, H_0 and H_1, and a decision criterion (Tanner & Swets, 1954).[2] H_0 and H_1 stand for two hypotheses, such as same or different, or signal or noise. The decision criterion balances the costs of

$$b = f(s,t)$$
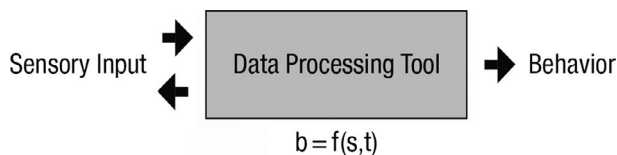
Fig. 2. Tools-to-theories. Behavior (*b*) is explained by a data processing tool (*t*) that analyzes the temporal and spatial structure of the sensory input (*s*). The tool is one that a community of researchers uses, such as a method of statistical inference or features of computer programs. The arrow pointing from the tool to the sensory input indicates that the tool determines the sensory input it can process, for instance, the data generated in an experiment (see text).

the two possible errors, false alarms (e.g., mistaking two equal tones as being different) and misses (mistaking two different tones as being identical). Finally, depending on what side of the criterion the sensory input (tone) falls, the decision is to classify the tones as the same or different. In this view, visual discrimination is a decision based on a balance of errors.

Signal detection theory was inspired by Neyman and Pearson's statistical decision theory and is formally identical to it. By the 1960s, statistical inference was institutionalized in experimental psychology, and the community became familiar with Type-1 and Type-2 errors, false alarms, and misses, as stated in the Neyman–Pearson theory. The tool required a new kind of data, hit rates and false alarm rates (Gigerenzer & Murray, 2015). Thus, unlike in earlier research in psychophysics and recall tasks in memory research, where only one kind of error was measured, that is, numbers of correct and incorrect responses, the new theory changed the data. This is indicated in Fig. 2 by the arrow from theory to input. It also provided a novel way to think about what is going on inside the black box and about research questions that had never been asked before. An example is the question of how the mind sets the decision criterion in order to balance the two possible errors, and what input-related factors determine the balance.

3.2. Example 2: Mind as computer

Herbert Simon and Allen Newell had argued since the late 1950s that higher-level cognition proceeds very much like a production system, a formalism from computer science. Yet the psychological community did not accept the analogy between cognition and computation before becoming familiar with the use of computers in the 1970s and 80s. Large-frame computers had been used before for data analysis and simulation, but were fraught with problems. The annual reports of the Harvard University Center for Cognitive Studies in the 1960s show that more time was spent on debugging and maintenance of their PDP-4C computer than on research, which meant that the computer was not yet an attractive research tool (Gigerenzer & Goldstein, 1996). Simon (1979) acknowledged that the acceptance of the mind as computer analogy was impeded by "the unfamiliarity of psychologists with computers" (p. 365). The wide acceptance of the idea that cognition is computation emerged only when desktop computers became available and affordable, and psychologists used them for data processing (Gigerenzer & Goldstein, 1996).

The tools-to-theories heuristic of discovery projected the formal structure of statistical tools onto theories of the mind, as illustrated by signal detection theory, but it also inspired new theories that used computational terms in more metaphorical ways. An example is Levelt's (1989) model of speaking, which postulates as its basic unit a "processing component" that corresponds to the computer programmer's concept of subroutine. Another example is Kelley's (1967) causal attribution theory, which postulates that the mind draws causal inferences by calculating an analysis of variance, a statistical tool that psychologists began to use excessively in the 1950s and 60s. Similarly, the development of nonmetric multidimensional scaling by Shepard (1962) inspired formal models of categorization, called exemplar theories, and the sequential decision theory of Wald

(1947) provided formal models for exemplar-random walk theories of categorization and for evidence accumulation models (e.g., Nosofsky & Palmeri, 1997; Ratcliff & Smith, 2004). At a more general level, cognitive developmental has been discussed as a change in "the *instruments* of scientific thinking" (Kuhn, 1989, emphasis in the original). For instance, in contrast to Jean Piaget's theories, "preschoolers test hypotheses against data and make causal inferences; they learn from statistics and informal experimentation" (Gopnik, 2012, p. 1623).

The tools-to-theories heuristic can lead to useful, novel, and general theories. But one should always take a close look at the assumptions the tool transfers to the mind. For instance, Levelt's model assumes the principle of isolated processing components, a maxim of computer programming, which has been criticized as unrealistic and as the Achilles' heel of the model (Gigerenzer & Goldstein, 1996).

Is tools-to-theories limited to cognitive psychology? It appears not. In personality psychology, for instance, factor analysis has inspired thinking of personality in terms of factors, such as the Big Five. In electrophysiology, the instruments sometimes suggested or even functioned as explanatory models for the phenomena, such as when electric current was used to analyze the working of nerves and muscles and the phenomena studied were then called "nerve current" and "muscle current" (Lenoir, 1986; see also Cowles, 2015). In a more controversial case, Koehn (2011) argued that financial theory, including Markowitz's Nobel Prize-winning mean-variance portfolio, was shaped by the analytical tools the theorists used, such as linear programming, a method to optimize two or more variables at one time (e.g., risk and return). Here, the tool created an entirely new discipline: finance theory. Yet the tool also cultivated the illusion in many theoreticians that the real world of finance is about risk rather than uncertainty, and that fine-tuned optimization (rather than robust heuristic methods; see below) can prevent further crises (Gigerenzer, 2018). As critics observed, even before the crisis of 2008 the theoreticians "seem[ed] to be more interested in demonstrating their mathematical prowess than in solving genuine problems" and had "lost virtually all contact with terra firma" (Durand, 1968, p. 848).

The tools-to-theories heuristic requires a re-evaluation of philosophical accounts of the relation between theory and data, and the presumed independence of the context of discovery from the context of justification (Gigerenzer & Sturm, 2007). In Popper's (1959) account, philosophy and statistics deal with the justification of theories, while the context of discovery "is irrelevant to the logical analysis of scientific knowledge" (p. 31). In this view, theories somehow emerge mystically, as illustrated by stories about Fechner, Kekulé, or Poincare that link discovery to beds, bicycles, and bathrooms. Real science begins only at the point of testing theories. According to a second account, the inductive view of scientific discovery, theories emerge rationally as empirical generalizations of data. In each of these two standard accounts, scientific tools have a neutral role. The tools-to-theories heuristic, in contrast, implies the existence of discoveries when new tools for data processing are introduced into a scientific community, from which new theories emerge that in turn require new kinds of data. Revising the view of science as a timeless, objective, and rational pursuit, Kuhn (1962) emphasized that observation and practice is

theory-laden. Yet that very insight made the study of practice appear to be of little relevance (Lenoir, 1988). In contrast, Hacking (1983) argued that experimental practice has a life of its own, and that theories can be practice-laden. When instruments change in the natural sciences, the theoretical argumentation is likely to change, too (Galison, 1987). In the same spirit, the tools-to-theories argument puts scientific practice back into philosophical accounts of discovery.

4. As-if explanations

One of the most influential explanations of behavior is the theory of expected utility maximization. It dominates neoclassical economic theory and has shaped psychological explanations of motivation (Atkinson, 1964), health behavior (Heckhausen, 1991), attitude formation (Fishbein & Ajzen, 1975), and moral behavior (Gigerenzer, 2010), among others. The theory addresses the question: How should one choose an action from a set of possible actions? To maximize one's expected utility, one needs to know the exhaustive and mutually exclusive set of actions along with their future consequences and probabilities. The expected utility of an action is defined as the sum of the products of the utilities and probabilities for each of its consequence. The action with the highest expected utility represents the rational choice. This basic theory has been modified in many ways, such as in prospect theory, and the learning of probabilities has been modeled by Bayes' rule.

Despite their widespread use, theories of expected utility maximization have been criticized as being computationally intractable, assuming perfect knowledge about the exhaustive set of actions and future consequences, and lacking in empirical evidence for the existence of stable utility functions (e.g., Friedman, Isaac, James, & Sunder, 2014). In his classic defense of expected utility theory, the economist Milton Friedman (1953) countered that the psychological realism of a theory does not matter. What matters is solely the theory's accuracy in predicting behavior. According to Friedman, an explanation is only *as-if*: People behave *as if* they maximized expected utility. What might be occurring in people's minds is not observable and thus of no relevance in science—an argument reminiscent of black-box behaviorism. Friedman's as-if argument is still the standard interpretation of expected utility models in economics today. Its implication is that cognitive psychology is largely irrelevant.

The classic example of an as-if theory is Ptolemy's theory of the movement of heavenly bodies, with the earth at the center of the universe and the sun and planets moving around it in circles. For a priori reasons, his model was geocentric and used circles to explain planetary movement. To make the theory work, small circles (*epicycles*) were added to the circles. This system led to excellent predictions, even though few may have believed that planets actually move in epicycles. In contrast, Kepler's theory aimed at describing the actual movement of planets, which required sacrificing the central position of the earth and the beauty of circles for ellipses. Kepler's theory remains a theory, but it describes the actual process—the movement of planets.

In psychology, the distinction between as-if and process models is often expressed in terms of Marr's (1982) distinction between a computational and an algorithmic level of explanation. An analysis at the computational level tries to understand the function of the cognitive system; an analysis at the algorithmic level tries to understand the cognitive process. Like an as-if theory, a computational analysis is mute about the nature of the cognitive process (Fig. 3).

4.1. Example 1: Bayesian brains

Inspired by economic theory, cognitive processes—including their neural underpinnings—have been modeled as Bayesian inference (Anderson, 1990). The central idea of this *rational analysis* program is that the mind makes inferences about the world by updating a prior probability distribution over the exhaustive set of possible actions into a posterior probability distribution, taking into account new evidence (the sensory input). The problem of Bayesian computations and their approximations being intractable is made irrelevant because there is no commitment to an algorithmic explanation. For instance, Tenenbaum and Griffiths (2001) clearly state: "We do not assert that any of our statistical calculations are directly implemented, either consciously or unconsciously, in human minds, but merely that they provide reasons why minds compute the way that they do" (p. 776). The aim of these theories is to describe how an ideal rational system, such as a Bayesian machine, would define and solve a problem, and to test whether people behave *as if* they performed these calculations.

4.2. Example 2: Egon Brunswik's intuitive statistician

Brunswik (1955) was probably the first to compare the mind to an intuitive statistician. He adopted the correlation methods used in personality and intelligence research, and analyzed perception in terms of linear multiple regression. Brunswik made it clear that he did not think that regression describes the cognitive processes, but that it is merely a "paramorphic" (his term for *as-if*) model for measuring cue validities and ecological validities. Brunswik's program can be seen as an antecedent of the rational analysis program.
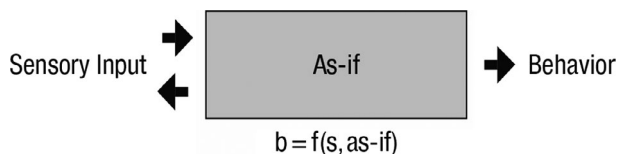

Fig. 3. As-if theories. Behavior (*b*) is explained by the maximization of expected utility, Bayesian probability updating, or some modification of these. The argument is not that these calculations are performed in the black box, but that people behave "as if" they performed them. The arrow pointing to the sensory input signals that the as-if theory in turn determines the sensory input it requires, for instance, the data provided to a participant in an experiment.

Although researchers do not always agree whether a theory is as-if or a process model, two distinguishing criteria exist. One was already mentioned: If the theory implies computations that are intractable, it is as-if. The second criterion is empirical: A process theory can be shown to be incorrect if it makes false predictions about either the process or the resulting behavior. An as-if theory, in contrast, is immune to critique of the reality of its assumptions about psychological processes and can only be tested by predictions about behavior. What complicates the matter is that one and the same theory may have parts that are as-if—say, because they are intractable—and others that are actually meant to model cognitive processes.

Like tools-to-theories, as-if theories tend to picture the human mind as a rational agent. It may be worth briefly pointing out the striking contrast between these two frameworks and the equally influential heuristics-and-biases program in cognitive psychology (Kahneman, 2011). In Kahneman's dual processing theory, statistical processes are aligned with conscious processes in System 2 that are considered rational, whereas intuitive processes are aligned with unconscious processes in System 1 that are said to lack rationality. Tools-to-theories, in contrast, describes the mind as unconsciously performing statistical calculations with no lack of rationality. Moreover, the use of formal models in both frameworks helps to overcome the limitations of a list of general dichotomies without formal precision, such as Systems 1 and 2.

As-if theories have not only had a considerable impact on the models of cognition, but also on society in general. The idea that people behave as if they maximized their expected utility is the fundament of neoclassical economics and the related neoliberal view that the state should intervene as little as possible in the market, business, health-care, and other human affairs. In this view, individuals and the market behave *as if* they had perfect foresight of the future consequences of their actions and updated probabilities consistent with Bayes' rule. In this view, governmental intervention can only disturb the equilibrium created by the "invisible hand."

5. Adaptive toolbox theories

Given that as-if theories describe a problem at a computational level, what then are the algorithms that the cognitive system uses to solve the problem? Proponents of as-if theories have argued that the actual cognitive processes are likely to consist of heuristics. One might ask, why not study the heuristics directly? Herbert Simon proposed exactly this solution in his critique of the almost universal use of as-if models in economics. He argued that researchers should (i) study how people actually make decisions instead of constructing as-if theories, and (ii) analyze how people make decisions under uncertainty, as opposed to risk (Knight, 1921; Savage, 1954). *Risk* refers to situations where the exhaustive and mutually exclusive set of future states of the world and all their consequences and probabilities are known for certain. For instance, roulette is a game of risk, where all possible future states of the world are known, namely the numbers 0 to 36; the full set of consequences, that is, the pay-offs; and their probabilities. Many psychological

studies present participants with problems of risk, such as having to choose between monetary gambles or to make moral choices in the trolley problems. *Uncertainty* refers to situations where this knowledge is not available, which Simon characterized as situations of *bounded rationality*. Some degree of uncertainty is inherent to virtually all important decisions in real life, such as what job offer to accept, where to invest money, and whom to marry. Under risk, probability theory is sufficient to determine the optimal solution of a problem. Under uncertainty, by contrast, an optimal solution cannot, by definition, be determined (Savage, 1954).

Simon proposed that under uncertainty, cognitive processes involve heuristics that can guide behavior quickly and accurately. Although Gestalt psychologists had used the term *heuristics* earlier to describe search for information, Simon insisted on formal models of heuristics and on studying how these are adapted to the structure of the environment. The adaptive use of heuristics became systematically studied in the work on the adaptive decision maker (Payne, Bettman, & Johnson, 1993) and on the adaptive toolbox (Gigerenzer, Todd, & the ABC Research Group, 1999). A typical model of a heuristic specifies rules for search, stopping search, and decision making. That is, it is a testable model of the order or direction in which information is searched, when search is stopped, and how the information is integrated into a decision or judgment. The term *adaptive toolbox* stands for the repertoire of heuristics an individual has learned, including the cognitive capacities required to execute these heuristics. In this view, behavior is a function of heuristics, which need to be adapted to the problem at hand. The arrow from the adaptive toolbox to the sensory input indicates that heuristics actively search and select the input that is processed in order to select the behavior (Fig. 4).

5.1. Example 1: Satisficing with aspiration-level adaptation

Lewin (1935) considered a successful person as one who sets goal values (aspiration levels) within reach. Simon (1955) applied this idea to situations where the assumptions required by expected utility maximization are not met. Here, people cannot optimize but have to satisfice. The satisficing heuristic consists of these steps: *Step 1: Set an aspiration level α. Step 2: Choose the first option that satisfies α. Step 3: If after time β no option has satisfied α, then change α by an amount γ and continue until an option is found.* Steps 1 and 2 define the basic satisficing model; Step 3 adds aspiration-level adaptation.
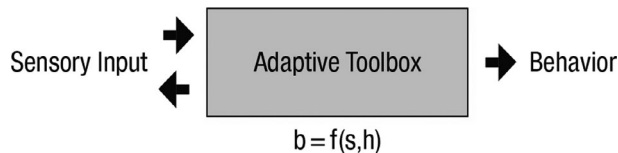

$$b = f(s,h)$$

Fig. 4. Adaptive toolbox. Behavior (*b*) is explained by a set of heuristics (*h*) that determine search through the sensory input (*s*). The set of heuristics available to an individual is called the *adaptive toolbox*. The arrow pointing from the adaptive toolbox towards the sensory input indicates that heuristics actively search and select the sensory input it can process.

The aspiration level can be a single attribute, or it may consist of multiple attributes. An example of multiple attributes x and y is the decision making of professional entrepreneurs who select investment options by this satisficing heuristic: If I expect at least x return within y years, then I take the option (Berg, 2014).

5.2. Example 2: Recognition heuristic

Recognition is a core cognitive capacity that the recognition heuristic exploits to make inferences under limited knowledge. In the case of the choice between two alternatives, the heuristic is: *If one of two objects is recognized and the other is not*, *then infer that the recognized object has the higher value with respect to the criterion*. This heuristic is ecologically rational in situations with high recognition validity. It makes a bold prediction that no other theory has made: the existence of less-is-more effects. For instance, people who know less about a topic can make systematically more accurate predictions in comparison with others who know more. This occurs when the recognition validity > knowledge validity; the conditions are specified in Goldstein and Gigerenzer (2002). The recognition heuristic can be generalized to choice between more than two alternatives, where it describes the creation of consideration sets.

Note that both examples are situations of uncertainty. A car dealer cannot know for sure what amount of money potential customers are willing to pay tomorrow, and participants in the laboratory experiment were told neither the validity of recognition and of other cues nor the ecological rationality of the heuristic. Under uncertainty, the optimal solution can, by definition, only be known in hindsight. This is why the algorithms that define the heuristics are not shortcuts or approximations of an optimal response. In fact, an analysis showed that the car dealers who relied on the satisficing heuristic earned more money than if they had relied on standard optimization models (Artinger & Gigerenzer, 2016). Under uncertainty, less fine-tuning and computation can lead to better decisions. Less can be more.

Models of heuristics specify the individual steps of the decision process, in addition to making predictions of the outcome (Gigerenzer, Hertwig, & Pachur, 2011). For instance, satisficing specifies a characteristic process: the setting and adjustment of aspiration levels. In the case of used car dealers, economic theories based on Bayesian probability updating assume that they set the price of a used BMW by constantly fine-tuning it to the latest market news. When studying the actual decision process of 628 used car dealers, we found that 97% of them instead relied on a satisficing heuristic to price their cars (Artinger & Gigerenzer, 2016). That is, they set an initial price, retained this price on average for about 4 weeks, and then lowered the price if the car was not sold, and so on.

A heuristic is adapted to specific environmental structure—hence the term *adaptive toolbox*. The match between a heuristic and an environment is the topic of the study of ecological rationality. As mentioned above, the recognition heuristic is ecologically valid if the recognition validity is substantially higher than chance. Studies show that people have an intuitive sense for situations in which the recognition heuristic is valid. For instance, name recognition of Swiss cities is a valid predictor of their population

(recognition validity = 0.86) but not of their distance from the center of Switzerland (recognition validity = 0.51, which is about chance). Accordingly, 89% of German students' inferences about which of two Swiss cities has the higher population followed the predictions of the recognition heuristic model but only 54% of their judgments followed its predictions when judging distance (Pohl, 2006). Across 43 experiments, the correlation between the percent of correct predictions (as a proxy for use of the heuristic) and recognition validity was $r = .57$. In general, the ecological rationality of a heuristic depends on the environmental structure, and people appear to be sensitive to this relation when choosing the heuristic.

Heuristics exploit existing cognitive capacities, such as memory. The recognition heuristic, for instance, has been implemented in the ACT-R model of memory (Schooler & Hertwig, 2005) and, alternatively, in a signal detection model of recognition memory (Pleskac, 2007).

The study of the mind as an adaptive toolbox follows Simon's requirement for formal models of cognitive processes to overcome the limits of research that uses ambiguous verbal labels for heuristics, such as *availability*. Formal models of heuristics can make specific predictions, such as when less-is-more effects occur, whereas labels can "explain" only after the fact due to their high flexibility. What the tools-to-theories framework and the study of formal theories of heuristic have in common is that they model cognitive processes. However, the study of heuristics does not assume that these processes correspond to the statistical tools used in a research community. For an introduction to the study of the mind's adaptive toolbox, see Gigerenzer et al. (2011).

6. On the relation between the three frameworks

The three frameworks can be related to each other by two dimensions. The first dimension is whether cognition is studied in situations of risk or uncertainty, and the second whether models of cognitive processes are constructed or not (Table 1).

I use the terms *risk* and *uncertainty* (Knight, 1921) as a shorthand for situations in which the probability calculus can deliver the optimal solution (risk) and in which it cannot (uncertainty). Playing the roulette is a situation of risk; hiring a professor or founding a start-up are situations of uncertainty. For the probability calculus to deliver the optimal solution, a "small world" of risk is required. Savage (1954), known as the originator of modern Bayesian decision theory, explicitly stated that Bayesian theory applies only to "small worlds"

Table 1
The three frameworks can be related to each other according to whether they study situations of risk versus uncertainty and whether they construct models of cognitive processes or not

	Risk	Uncertainty
Models of Cognitive Process	Tools-to-Theories	Adaptive Toolbox Theories (Heuristics)
No Models of Cognitive Process	As-if Theories	

and not beyond. A small world {S, C}, or a situation of risk, consists of the exhaustive and mutually exclusive set S of future states of the world and the exhaustive and mutually exclusive set C of consequences. Savage noted that it would be "utterly ridiculous" to apply it to situations like "planning a picnic" or "playing chess" (p. 16). Planning a picnic represents ill-defined situations where one cannot know S and/or C because certain and unexpected events may happen, while chess is a well-defined problem that is intractable.

The distinction between risk and uncertainty helps to clarify the relation between tools-to-theories and heuristics. Most tools that have been reconsidered as theories of mind—including Fisher's ANOVA, Neyman–Pearson statistics, and Wald's sequential statistics—are optimization tools. They apply to situations of risk, whereas heuristics address uncertainty. For instance, when signal-detection theory, which is an optimization theory, is applied to out-of-sample prediction, which represents a minimal form of uncertainty, simple heuristics can predict more accurately and with less effort than the optimization model (Luan, Schooler, & Gigerenzer, 2011, 2014). The distinction between risk and uncertainty also indicates the irreducible value of models of heuristics in order to extend theories of cognition to the many real-world situations that involve uncertainty.

What is the relationship between theories that model cognitive processes, models of heuristics and tools-to-theories, and as-if theories? As mentioned before, neo-classical economics uses expected utility maximization and Bayesian probability updating as its universal framework—and often as a prerequisite for publication in leading journals. Like radical behaviorism, it has little to no interest in an analysis of psychological processes. Moreover, theories in behavioral economics such as prospect theory and hyperbolic discounting add free parameters to the expected utility calculus, thereby producing more complex as-if models (Berg & Gigerenzer, 2010). Similarly, proponents of Bayesian models of cognition largely restrict themselves to building as-if models, arguing that these correspond to Marr's computational level. For instance, Xu and Tennenbaum (2007, pp. 250–251) write that their "framework aims to explain inductive learning at the level of *computational theory* (Marr, 1982) . . . rather than to describe precisely the psychological processes involved" (italics in original). In this interpretation of Marr's view, heuristics are to Bayesian rationality what the actual algorithms in a pocket calculator are to the theory of arithmetic. Because people's rationality is bounded by time and memory, so the argument goes, people may rely on "approximate" algorithms (heuristics) to reach a goal (e.g., Chater & Oaksford, 1999). With few exceptions, no attempt is made to model the cognitive processes, resulting in a program that might be called "Bayesian behaviorism." Thus, there are two related theses: (a) Bayesian statistics provides the optimal theoretical solution, whereas the study of cognitive processes only shows how people approximate optimality given their cognitive constraints, which justifies that (b) the study of cognitive processes is of little relevance.

Thesis 1, however, is incorrect as a universal statement about cognition, and Thesis 2 therefore does not follow. Because Thesis 1 has been justified by reference to Marr's levels, let me begin with what Marr really wrote. First, the claim that Bayesian statistics is a universal computational theory is not Marr's. In fact, Marr (1982) does not even mention Bayes. Second, and more important, the claim that there is a computational theory for all problems,

Bayesian or otherwise, was rejected by Marr himself (see Brighton, in press). Marr (1977) distinguished between "Type 1" theories that yield to a computational analysis and "Type 2" theories that do not (such as the grammar of natural languages). According to Marr, Type 2 theories are the majority in artificial intelligence.

The distinction between risk and uncertainty makes it clear that the relation between as-if theories and models of heuristics is *not* identical to that between the computational and the algorithmic level (Thesis 1). Under uncertainty, optimal solutions by definition do not exist (Marr's Type 2 theories); it therefore makes little sense to believe that heuristics are inferior because they do not offer optimal solutions (Thesis 1) and that studying cognitive processes is of little relevance (Thesis 2). In other words, both Theses 1 and 2 are misleading for situations that involve uncertainty. Orthodox responses in contending with uncertainty have been uninformed priors, imprecise probabilities, and second-order probabilities. These, however, can deal solely with ambiguity (unknown probabilities), not with a situation of uncertainty that has an unknown state space {S, C}. Under uncertainty, a rational analysis that determines the optimal solution is by definition a fiction, but can be replaced by the study of the ecological rationality of heuristics, that is, the environmental conditions under which heuristics succeed or fail relative to more complex strategies (Gigerenzer, Hertwig, & Pachur, 2011). Therefore, heuristics are not humans' approximations to optimality; they are cognitive tools when optimality is out of reach, as is the case in most real-world situations (Brighton, in press).

The scheme in Table 1 has its limits. There are overlaps, such as heuristics for decision making under risk (e.g., Brandstätter, Gigerenzer, & Hertwig, 2006), as well as theories that are hybrids of as-if models and process models. These hybrids could be placed in the empty cell in the table if they give up the ideal of optimality and actually model situations under uncertainty (e.g., Tauber, Navarro, Perfors, & Steyvers, 2017).

There is also another way to look at the differences between the three frameworks. Theories of cognition have always been inspired by analogies such as wax tablets, holograms, and dictionaries (Roediger, 1980). Tools-to-theories has been inspired by how researchers analyze data, as-if theories by how rational choice theorists think, and the study of the adaptive toolbox by how experts make decisions under uncertainty. In sum, there is a clear division of labor between the three approaches. Formal models of heuristics are the choice in situations of uncertainty, and the study of their ecological rationality is the answer to the question of under which conditions a given heuristic is likely to succeed, according to a criterion. As-if theories assuming expected utility maximization and Bayesian updating apply instead to situations of risk, where they can provide a normative benchmark but without insight into the cognitive processes. Tools-to-theories are typically meant for situations of risk but, unlike as-if theories, aim at providing insight into the cognitive processes.

7. Frameworks that determine researchers' questions

In this article, I distinguish and briefly describe three kinds of explanations that emerged after the cognitive revolution. Together, these are embodied in many present-

day theories of cognition. Yet these explanations are not identical, and they portray the content of the black box in systematically different ways. Nevertheless, they share a common value: the use of formal models in place of merely verbal concepts.

My analysis ends with open questions. First, little research exists on the scope of tools-to-theories explanations and on how exactly properties of the tool have shaped theories of cognition. In fact, the tool-based origin of theories is rarely pointed out. For instance, in their introduction to signal detection theory, Tanner and Swets (1954) mention that visual detection is "the task of testing a statistical hypothesis" (p. 403), and in other work, they compare the ideal observer to a "Neyman-Pearson detector" (Gigerenzer & Murray, 1987/2015, p. 52). Yet subsequent publications on signal detection theory have tended to ignore its origin. Similarly, articles presenting evidence accumulation and sequential sampling models rarely mention their origin in Wald's sequential decision theory, and thus there is little analysis of the assumptions that have been carried over into theories of memory and categorization. The tools-to-theories heuristic, by contrast, ascribes research tools a more prominent role than they have in psychology and in the philosophy of science. A new tool for data processing can inspire a new theory of cognition, and this theory in turn can inspire new kinds of data.

Second, although economists have explicitly promoted as-if explanations of expected utility and Bayesian updating, psychologists do not always distinguish whether a theory is meant to be as-if or a model of the process. The major exception is when Marr's distinction between computational and algorithmic explanations is invoked. As I have argued, however, and as emphasized by Marr (1977) himself, that distinction is irrelevant when cognition has to deal with uncertainty that cannot be measured by probability. In addition, it is sometimes suggested that a computational theory poses constraints for the heuristic processes that a cognitive system uses, but whether this is true has not been demonstrated and remains controversial (Brighton & Gigerenzer, 2008).

Finally, as-if models are motivated by the belief that a single mathematical tool such as utility maximization or Bayes' rule could serve as a universal theory for all behavior, just as Bayesian statistics—equally wrongly, in my opinion—is considered a universal method of scientific inference (Colombo, Elkin, & Hartmann, 2018; Gigerenzer & Marewski, 2015). The alternative to this universal inference procedure in the brain or in science is a toolbox approach, as in the statistical toolbox that generates tools-to-theories and in the adaptive toolbox of heuristics. The distinction between risk and uncertainty is one means of understanding that the belief in a Bayesian brain or a universal method for scientific inference is a beautiful illusion. If the mind were Bayesian alone, it could hardly survive. It could not conceive of new ideas, would suffer from overfitting when making predictions with scarce data but many variables, and would be lost when faced with intractable problems.

I think that psychology needs two agendas that editors of journals should consider promoting. First, we need to think more about the limits of theories. No theory should be published without specification of the domain where it does not work or apply. The program of ecological rationality is an example of such a program, which states the conditions under which a given heuristic will succeed or fail (Gigerenzer et al., 2011). Second,

we need to think more about the integration of existing theories. Theory integration is an alternative and complementary route to Popper's program of theory elimination and, in my opinion, one of the most vital challenges to strengthening the theoretical fundament of psychology. I have outlined such a program (Gigerenzer, 2017), which aims at connecting existing theories by showing that apparently disparate phenomena or concepts are theoretically connected. Each of the three frameworks distinguished in this article has limits in terms of what it can explain, such as process or outcome, uncertainty, or risk. Yet these frameworks also have the potential to inform each other and to lead to points of integration, a potential that is open for exploration.

Notes

1. This account refers to psychology in the United States; in Europe, behaviorism never dominated psychology in the first place.
2. This account is consistent with Tanner's original interpretation of the theory (see Gigerenzer & Murray, 2015, pp. 49–53). Alternatively, one could assume that the decision maker starts with some more or less arbitrary criterion and adapts it to the samples encountered.

References

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

Artinger, F. M., & Gigerenzer, G. (2016). The cheap twin: From the ecological rationality of heuristic pricing to the aggreagate market. *Academy of Management Proceedings*, *2016*(1), 13915. https://doi.org/10.5465/ambpp.2016.206

Atkinson, J. W. (1964). *An introduction to motivation*. Princeton, NJ: Van Nostrand.

Berg, N. (2014). Success from satisficing and imitation: Entrepreneurs' location choice and implications of heuristics for local economic development. *Journal of Business Research*, *67*, 1700–1709. https://doi.org/10.1016/j.jbusres.2014.02.016

Berg, N., & Gigerenzer, G. (2010). As-if behavioral economics: Neoclassical economics in disguise? *History of Economic Ideas*, *18*, 133–165. https://doi.org/10.1400/140334

Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2006). The priority heuristic: Making choices without trade-offs. *Psychological Review*, *113*, 409–432.

Brighton, H. (in press). Rationality without optimality: Bounded and ecological rationality from a Marrian perspective. In R. Viale (Ed.), *Routledge handbook of bounded rationality*. London: Routledge.

Brighton, H., & Gigerenzer, G. (2008). Bayesian brains and cognitive mechanisms: Harmony or dissonance? In N. Chater & M. Oaksford (Eds.), *The probabilistic mind: Prospects for Bayesian cognitive science* (pp. 189–208). New York: Oxford University Press.

Brunswik, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*, 193–217. https://doi.org/10.1037/h0047470

Chater, N., & Oaksford, M. (1999). Ten years of the rational analysis of cognition. *TRENDS in Cognitive Sciences*, *3*, 57–65. https://doi.org/10.1016/S1364-6613(98)01273-X.

Chomsky, N. (1959). Review of Skinner's verbal behavior. *Language Sciences*, *35*, 26–58.

Colombo, M., Elkin, L., & Hartmann, S. (2018). Being realist about Bayes, and the predictive processing theory of mind. *The British Journal for Philosophy of Science*, axy059. https://doi.org/10.1093/bjps/axy059

Cowles, H. M. (2015). Hypothesis bound: Trial and error in the nineteenth century. *Isis*, *106*, 635–645.

Durand, D. (1968). State of the finance field: Further comment. *The Journal of Finance*, *23*, 848–852.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Reading, MA: Addison-Wesley.

Friedman, D., Isaac, R. M., James, D., & Sunder, S. (2014). *Risky curves. On the empirical failure of expected utility*. New York: Routledge.

Friedman, M. (1953). *Essays in positive economics*. Chicago, IL: University of Chicago Press.

Galison, P. (1987). *How experiments end*. Chicago, IL: University of Chicago Press.

Garcia, J., & Koelling, R. A. (1966). The relation of cue to consequence in avoidance learning. *Psychonomic Science*, *4*, 123–124.

Gigerenzer, G. (1991). From tools to theories: A heuristic of discovery in cognitive psychology. *Psychological Review*, *98*, 254–267. https://doi.org/10.1037/0033-295X.98.2.254

Gigerenzer, G. (2010). Moral satisficing. Rethinking moral behavior as bounded rationality. *Topics in Cognitive Science*, *2*, 528–554. https://doi.org/10.1111/j.1756-8765.2010.01094.x

Gigerenzer, G. (2017). A theory integration program. *Decision*, *4*, 133–145. https://doi.org/10.1037/0033-295X.98.2.254

Gigerenzer, G. (2018). The heuristics revolution: Rethinking the role of uncertainty in finance. In R. Viale, S. Mousavi, B. Alemanni, & U. Filotto (Eds.), *The behavioral finance revolution: A new approach to financial policies and regulations* (pp. 115–134). Cheltenham, UK: Edward Elgar.

Gigerenzer, G., & Goldstein, D. G. (1996). Mind as computer: Birth of a metaphor. *Creativity Research Journal*, *9*, 131–144.

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

Gigerenzer, G., & Marewski, J. (2015). Surrogate science: The idol of a universal method for scientific inference. *Journal of Management*, *41*, 421–440. https://doi.org/10.1177/0149206314547522.

Gigerenzer, G., & Murray, D. J. (2015). *Cognition as intuitive statistics* (Psychology Revivals Series). London: Psychology Press. (Original work published 1987).

Gigerenzer, G., & Sturm, T. (2007). Tools = theories = data? On some circular dynamics in cognitive science. In M. G. Ash & T. Sturm (Eds.), *Psychology's territories: Historical and contemporary perspectives from different disciplines* (pp. 305–342). Mahwah, NJ: Erlbaum.

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

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, *109*, 75–90. https://doi.org/10.1037/0033-295X.109.1.75

Gopnik, A. (2012). Scientific thinking in young children: Theoretical advances, empirical research, and policy implications. *Science*, *337*, 1623–1627. https://doi.org/10.1126/science.1223416

Hacking, I. (1983). *Representing and intervening*. Cambridge, UK: Cambridge University Press.

Heckhausen, H. (1991). *Motivation and action* (P. K. Leppmann, Trans.). Berlin: Springer-Verlag.

Kahneman, D. (2011). *Thinking, fast and slow*. London: Allen Lane.

Kelley, H. H. (1967). Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska symposium on motivation*. (vol. *15*, pp. 192–238). Lincoln: University of Nebraska Press.

Knight, F. (1921). *Risk, uncertainty and profit* (vol. *XXXI*). Boston, MA: Houghton Mifflin.

Koehn, J. (2011). From tools to theories: The emergence of modern financial economics. Discussion Paper No. 16/2011, University of Witten/Herdecke, Faculty for Economics. Retrieved from http://hdl.handle.net10419/52235.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review*, *96*, 674–689. https://doi.org/10.1037/0033-295X.96.4.674

Kuhn, T. (1962). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.

Lenoir, T. (1986). Models and instruments in the development of electrophysiology, 1845–1912. *Historical Studies in the Physical Sciences*, *17*, 1–54. https://doi.org/10.2307/27757574

Lenoir, T. (1988). Practice, reason, context: The dialogue between theory and experiment. *Science in Context*, *2*, 3–22. https://doi.org/10.1017/S0269889700000478

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Lewin, K. A. (1935). *A dynamic theory of personality*. New York: McGraw-Hill.

Luan, S., Schooler, L., & Gigerenzer, G. (2011). A signal detection analysis of fast-and-frugal trees. *Psychological Review*, *118*, 316–338. https://doi.org/10.1037/a0022684

Luan, S., Schooler, L. J., & Gigerenzer, G. (2014). From perception to preference and on to inference: An approach-avoidance analysis of thresholds. *Psychological Review*, *121*(3), 501–525. https://doi.org/10.1037/a0037025

Marr, D. (1977). Artificial intelligence: A personal view. *Artificial Intelligence*, *9*, 37–48. https://doi.org/10.1016/0004-3702(77)90013-3

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco, CA: W. H. Freeman.

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, *104*, 266–300.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge, UK: Cambridge University Press.

Pleskac, T. J. (2007). A signal detection analysis of the recognition process. *Psychonomic Bulletin & Review*, *14*, 379–391.

Pohl, R. F. (2006). Empirical tests of the recognition heuristic. *Journal of Behavioral Decision Making*, *19*, 251–271. https://doi.org/10.1002/bdm.522

Popper, K. (1959). *The logic of scientific discovery*. New York: Basic Books.

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, *111*, 333–367. https://doi.org/10.1037/0033-295X.111.2.333

Roediger, H. L. (1980). Memory metaphors in cognitive psychology. *Memory & Cognition*, *8*, 231–246.

Savage, L. J. (1954). *The foundations of statistics*. New York: Wiley.

Schooler, L. J., & Hertwig, R. (2005). How forgetting aids heuristic inference. *Psychological Review*, *112*, 610–322. https://doi.org/10.1037/0033-295X.112.3.610

Shepard, R. N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function. Part 2. *Psychometrika*, *27*, 219–246. https://doi.org/10.1007/BF02289630

Simon, H. A. (1955). A behavioral model of rational choice. *Quarterly Journal of Economics*, *69*, 99–118. https://doi.org/10.2307/1884852

Simon, H. A. (1979). Information processing models of cognition. *Annual Review of Psychology*, *30*, 363–396. https://doi.org/10.1146/annurev.ps.30.020179.002051

Tanner Jr., W. P., & Swets, J. A. (1954). A decision-making theory of visual detection. *Psychological Review*, *61*, 401–409.

Tauber, S., Navarro, D. J., Perfors, A., & Steyvers, M. (2017). Bayesian models of cognition revisited: Setting optimality aside and letting data drive psychological theory. *Psychological Review*, *124*, 410–441. https://doi.org/10.1037/rev000005

Tenenbaum, J. B., & Griffiths, T. L. (2001). Some specifics about generalization. *Behavioral and Brain Sciences*, *24*, 762–778. https://doi.org/10.1017/S0140525X01780089

Wald, A. (1947). *Sequential analysis*. New York: Wiley.

Xu, F., & Tennenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review*, *114*, 245–272. https://doi.org/10.1037/0033-294X.114.2.245