A quantum information architecture for cue-based heuristics

Peter D. Kvam Max Planck Institute for Human Development Michigan State University Timothy J. Pleskac Max Planck Institute for Human Development

Please not that this document is a preprint, and may not reflect exactly the content of the published article. Please refer to the official version on the APA site once it is available. http://dx.doi.org/10.1037/dec0000070

Cues indicating the state of the world play a critical role in decision-making in both inferential and preferential tasks, and are the focus of many heuristic models of cognitive processes. In this paper, we present the formal logical structure that is presupposed in a specific class of these processes: fast and frugal heuristics. We review the structure of these heuristics and show that they make a number of implicit assumptions resulting from the formal classical logic that underpins their structure. These assumptions lead to a number of predictions that are inconsistent with existing empirical data, but they follow primarily from the classical logic structure of heuristics and not from the rule-based cue processing approach itself. We introduce an alternative logic based on quantum theory that proposes a different and potentially more accurate set of underlying assumptions, and show that this framework addresses many of the issues arising from classical logic. We then demonstrate how heuristics can be reconstructed by integrating them with a quantum logic structure, examine the benefits that heuristics and quantum logic provide to one another, and outline the new questions and predictions that their integration yields. We contend that the integration of heuristics with quantum logic enhances both frameworks, improving heuristics as descriptive models by bringing them closer to the empirical data and grounding quantum logic in a psychological theory by giving it concrete processing rules to implement.

Keywords: information theory, logic, heuristics, quantum cognition, decision-making

Introduction

The use of *cues* – indicators of the state of the world present in our environment – is an integral component of inferential and preferential decision making. Cues are often external indicators such as linguistic cues that differentiate words (Reed, 1989), visual cues that indicate distance (Jacobs, 1999), or nonverbal signals of deceptive behavior (De-Paulo et al., 2003), but it is also possible to use internal states such as fatigue, activated schemas, or working memory content (e.g., Koriat, 1997). The question that every decisionmaker must address is which cues to use in which situations. One solution to this problem is to establish rules for informa-

tion processing that specify general rules of cue organization that can be applied across multiple contexts. In the first section of this paper, we review some of the rules that people appear to use for common decisions. For simplicity, we focus on fast and frugal heuristics. These strategies use finite sets of cues to make primarily binary decisions (Gigerenzer et al., 2011, 1999; Hertwig et al., 2013; Todd et al., 2012). To do so, they propose a particular set and order of cue use, establishing a computationally tractable basis for information processing in decision tasks. In this paper, our examples center primarily around the recognition heuristic (Goldstein & Gigerenzer, 2002), in part because it has been well-studied and has proven to be empirically supported in a number of situations (see Pachur et al., 2011; Marewski et al., 2010, for reviews) but also because it has a straightforward logical structure that is simple enough to deconstruct. We also show how classical and quantum information processing architectures can be used to model many non-compensatory heuristics like take-the-best (Gigerenzer & Goldstein, 1996), and we present a brief outlook on how to construct other types of heuristics that involve summing and weighting or parallel cue processing as well.

In press at Decision, Special Issue on Theory Integration

Address correspondence to Peter Kvam, Department of Psychology, Michigan State University, East Lansing, MI 48824; E-mail: kvam.peter@gmail.com. We express our gratitude to Jerome Busemeyer, Laura Martignon, and Gerd Gigerenzer for their constructive comments on this work. The first author was supported by a graduate fellowship from the National Science Foundation (DGE-1424871) during development and writing of this paper.

An important insight that we gained from this work and we hope to convey here is that when establishing rules for information processing, one commits to a particular theory of information and information processing. In the case of heuristics, the information processing theory that has been adopted is a classical one, applying logic operators to deterministic bits of information. This is indicated by their ability to be diagrammed in information processing steps (i.e., box and arrow diagrams) (Gigerenzer & Goldstein, 1996; Gigerenzer et al., 1999), fast-and-frugal trees (Martignon et al., 2003, 2008; Luan et al., 2011), and implemented by deterministic rule systems like ACT-R (Schooler & Hertwig, 2005; Marewski & Schooler, 2011). Adopting a theory of information and information processing has consequences for a model's ability to describe human decision making. In the second section of the paper, we examine the basic principles of classical information theory and their implications for heuristics and other descriptive models of cognition.

Following our examination of classical logic heuristics, we introduce an alternative framework for constructing cue processing models based on quantum logic (Busemeyer & Bruza, 2012; Yanofsky et al., 2008; Nielsen & Chuang, 2010). We do this for two primary reasons. First, the quantum approach is contrasted with the classical one in order to elucidate the assumptions that accompany the use of classical information theory in models of cognition. In doing so, we also review empirical evidence for violations of the assumptions that are made by these models and their consequences for modeling heuristic processes. Second, we examine how the quantum logic framework could be used to construct heuristic models of cue processing. We show that doing so has a number of important consequences, focusing on stochastic and continuous representations of cues and beliefs, rule combination and parallel processing, and cuecriterion belief entanglement.

We aim to show that ultimately, each of the theories is enhanced by the integration. Quantum logic implementations of heuristics inherit many benefits by basing their structure on simple heuristics. At the same time, heuristics can gain empirical accuracy and make new predictions by being implemented in a quantum logic framework. Each of the constituent theories brings important features to the integration. Heuristics provide the structure and rules to begin constructing strategies - such as the search, stopping, and decision rules that a decision-maker implements (Gigerenzer, 2004). They also organize cues in systematic ways and serve as a computationally simple basis which can be used as-is to make straightforward models. In turn, quantum logic provides a more general framework for making cue processing rules, allowing for more flexible (but potentially more computationally complex) strategies that can enhance the descriptive and explanatory power of heuristics as models of human decision behavior.

It is worth noting at this point that our interest is primarily in developing better descriptive models of human behavior. Heuristics have a number of additional advantages in that they are easy to communicate and follow, which allows them to be well-utilized as prescribed strategies or models for decision procedures (i.e. what a person *should* do in a particular situation). However, our integration of heuristics with quantum logic is not intended to imply that these prescriptive strategies should also be modified. We hope that the integration results in a better account of how people actually make decisions, but do not intend to stake any claims regarding what we should tell them to do in terms of strategies.

An important feature of theory integration that we seek to illustrate here is that putting theories together results in novel predictions. We show that our integration results in new questions and hypotheses regarding the number of cues used, choice proportions, and the effects of measuring or changing people's beliefs about cues and criteria. In order to address the new predictions that arise from combining heuristic and quantum approaches, we examine results from work on recognition memory, perception, game theory, expertise, and the hindsight bias in decision-making. We conclude by examining possible extensions and limitations, related concepts of uncertainty, and the overall benefits of integrating quantum and heuristic approaches.

Heuristics

We define a *heuristic* as a decision strategy that ignores some amount of relevant information, usually with the goal of achieving faster, more accurate, or more frugal decisions. This definition is similar to the one proposed by Gigerenzer & Gaissmaier (2011), though other authors and programs of research define them in different ways (see Kahneman & Frederick, 2002; Shah & Oppenheimer, 2008; Simon, 1990). This definition encompasses a wide range of potential strategies, but common among them is that they specify what cues should be used and how different sets should be mapped onto decisions. Many heuristics are based on the Brunswikian lens model (Brunswik, 1956), which posits that objects in the environment about which we must make inferences give off indicators of their state or value [criterion], such as whether they are a predator or prey, edible or inedible, or larger or smaller. These indicators are in turn used by a decisionmaker based on how well they diagnose the state of the world, referred to as the ecological validity of the cue. Formally, the validity of a cue v(c) is defined as

$$v(c) = \frac{\text{\# times cue yields the correct inference}}{\text{\# times yields an inference}}.$$
 (1)

Validity is integral in a number of heuristics, including the take-the-best strategy (Gigerenzer & Goldstein, 1996; Gigerenzer et al., 1999). This heuristic ranks cues by their validity and checks them from highest to lowest validity until one discriminates between decision alternatives to yield an inference about the criterion. This presents a reliable route to a decision – expected accuracy can be approximated by the validity of the most diagnostic available cue, and the probability of reaching a diagnosis approaches one as the number of (non-perfectly correlated) cues inspected increases. Moreover, Gigerenzer & Goldstein (1996) showed that using such a heuristic can lead to performance as good or better than other rules and models that use all the cues, such as with a weighted additive decision rule (see also Martignon & Hoffrage, 2002).

Similarly, cue use can be constrained simply by their physical or psychological accessibility. For example, the ability to recognize a first alternative and not the second prevents a decision maker from immediately accessing other cues about the second alternative. In essence, no other cues can be retrieved because no other cue values are known. However, the lack of recognition for one alternative allows for use of the recognition heuristic (Goldstein & Gigerenzer, 2002; Pachur & Hertwig, 2006; Pohl, 2006), where one alternative is chosen if it is recognized and the other is not. As one might expect, this strategy is most effective when recognition is indicative of the actual value of the criterion. However, its performance also depends on how frequently it can be used. The recognition heuristic is most effective when the number of alternatives that are recognized is similar to the number that are not recognized - this situation will generate the greatest number of item pairs where one item is recognized and the other is not, so that recognition can be used as a discriminating cue. As a result, recognizing a limited number of alternatives can be beneficial by virtue of increasing the frequency with which the heuristic can be used, referred to as a less-is-more effect (Goldstein & Gigerenzer, 2002; Pachur et al., 2011).

Heuristic approaches to decision-making offer a number of advantages. Most importantly, they provide preset strategies that can be applied to new situations and interchanged for different environmental conditions, providing a decisionmaker with desirable (if not optimal) outcomes. They are also implementable by limited-capacity agents, and use a limited set of cognitive operations that could be combined and ordered in order to create new strategies. This makes them easily communicable to other agents, meaning that they can be developed or learned directly instead of relying on the comparatively slow modification of behavior-relevant genes via natural selection (see also Hutchinson & Gigerenzer, 2005). From a modeling point of view, the simplicity and set operations are also convenient, as they can be constructed using a limited set of quantifiable rules and procedural logic.

Another major advantage of heuristic strategies is that they do not need to perform the computations required to estimate the full covariance structure of the available cues. Because each cue is generally treated independently, these strategies do not need to consider whether or to what degree different cues are redundant with one another. Therefore, they often need much less data in order to generate a successful approach, and can be used quite early in the learning process to gain desirable outcomes. This may make them more likely to be used to computationally limited agents (humans), as environmental and selection pressures might favor these straightforward decision strategies.

Similarly, heuristics are computationally simple in terms of fitting and predicting data. Because they use few (if any) free parameters, there is little uncertainty in terms of estimation. In many cases, this may make them parsimonious explanations of human decision behavior, especially when limited choice data is available.

As with most models, heuristics have some limitations in terms of their descriptive accuracy (see also Dougherty et al., 2008; Gigerenzer et al., 2008, for issues regarding cue validity and organization). One important feature is the deterministic nature of these heuristics. While this makes strategies computationally simple, a person using a particular heuristic will always reach the same answer for a given pair of alternatives. As such, they can miss out on explaining the variability we observe in individuals' and groups' behavior.

Some variability in terms of choice predictions can be introduced by including "trembling-hand" errors - difference choices which occur simply because of some mistake by the decision-maker. However, even this is not enough to explain deviations from deterministic behavior. For example, Newell et al. (2003) found that by applying a criterion of 10% errors, only about 30% of participants searched for information, stopped searching, and decided in a manner consistent with take-the-best (i.e., a deterministic heuristic). One harsh interpretation of this result is that relatively few people actually use take-the-best. However, an alternative possibility is that a deterministic model of take-the-best does not fully describe how people use the heuristic and that what is needed is a stochastic model of take-the-best (see also Bröder & Schiffer, 2003; Lee & Cummins, 2004; Newell, 2005; Pohl, 2006; Glöckner & Bröder, 2011; Davis-Stober & Brown, 2011). As we show later in this paper, a quantum logic architecture of heuristics provides a principled means to construct just such a stochastic model of heuristics like take-the-best.

Furthermore, it seems likely that selection pressures in our evolutionary history would not have favored fully deterministic strategies in all situations. This is largely because they can be exploited in multi-agent interactions – if opponents are aware of the rules a decision-maker is using, they can tailor counter-strategies based on what they know the decision-maker will do. For example, take the matching pennies game: in this interaction, each of two players places a penny heads-up or tails-up. Player A receives a reward if the players' pennies show different sides and a penalty if they show the same sides, whereas player B receives a reward if the pennies show the same sides and a penalty if they show different ones. ¹ The optimal behavior in this game is to choose stochastically, 50% heads / 50% tails. If player A chooses deterministically based on the time of day, surroundings, or some other factor, player B may be able to pick up on these and guarantee a win.

In the face of these challenges, it seems selection pressures might have resulted in the evolution of a mechanism that can generate response variability and unpredictability. This matches well with the great variability that we observe in human behavior (as well as its unpredictability, to the dismay of a great many cognitive modelers). Therefore, a successful integration might endow heuristics with a mechanism for generating random or at least variable behavior, in turn bringing them closer as a description of human behavior.

Another facet behavior that is perhaps difficult to capture with classical heuristics is that humans can sometimes use cues that are further down the hierarchy than more valid, available cues. For example, take-the-best and the priority heuristic (Brandstätter et al., 2006; Gigerenzer et al., 1991) both posit a specific order of cue inspection, and they are quickly falsified by instances where more cues are considered (Glöckner & Betsch, 2008; Newell & Shanks, 2003; Pachur et al., 2008). This issue appears largely due to the strict serial processing hierarchy of the heuristics. Similarly, the strict serial processing that some heuristic models posit also struggle with cases of parallel cue processing, despite the largely parallel nature of information processing in the brain (McLeod et al., 1998). They might be brought closer to observed decision behavior, then, by allowing some mechanism for bypassing the most valid cues or for processing multiple cues at the same time. We cover one such method in our integration.

Finally, heuristics can be limited in their descriptive power by the binary nature of the cues they use. Even when an underlying construct like recognition seems to be continuous or at least multiple-valued, a threshold is typically applied to sort it into one category or another (Brandstätter et al., 2006). While the ability to adjust the cutoff can be a particularly beneficial adaptation (Pleskac, 2007; Luan et al., 2011), applying it can throw away valuable information regarding the source or object of interest. Similarly, disregarding cue values that are uncertain (e.g. partially recognized or somewhat believed to be present) by skipping them (Gigerenzer et al., 1991) or assuming them to be a particular value when unknown (Gigerenzer & Goldstein, 1996) prevents this uncertainty from being beneficially leveraged in decisions.

It is important to note that not all of these characteristics are ubiquitous across all programs and heuristics – for instance, continuous cues and parallel processing can be deliberately implemented (Hogarth & Karelaia, 2007). But the tendency of heuristics to possess deterministic, serial, and binary properties is no coincidence. As we illustrate in the next section, they arise because heuristics in their current form rely on a classical logic framework. This need not be the case. Instead, heuristics can be reconstructed using alternative logical frameworks that address many of these issues. In later sections, we show that quantum logic offers an attractive alternative that produces probabilistic choice behavior, shows how cue processing can unfold in and transition between serial and parallel, and represents uncertainty in cues and beliefs such that it can exploit uncertainty to achieve potentially better performance. At the same time, fast and frugal heuristics provide important theoretical guidance to a quantum logic architecture in terms of modeling human cognition. As we will show, heuristics provide psychologically plausible rules, computational simplicity, and efficiency to an architecture that can easily drift away from these properties.

Information processing

Particularly in the early days of computing, information theory was deliberately used as a formal basis for models of cognitive processes, including perception and schema activation (Axelrod, 1973), conceptual reasoning (Sowa, 1983), perceptual discrimination, and working memory (Beebe-Center et al., 1955; Garner, 1953; Miller, 1956; Pollack, 1952). While many current theories of cognition follow in the same tradition of information processing, few modern ones explicitly utilize the formal components of information theory.

The formal mathematical structure underlying information processing uses Boolean logic and algebra, transforming bits of information conditional on the inputs and structure of logical operators [gates]. Doing so allows a modeler to reduce many theories of cognitive processes to a formal mathematical form, and elucidates some of the basic assumptions of models or theories that are based on classical logic. In turn, the principles and implications of classical logic can be tested so that we might judge their suitability as a basis for modeling cognition. To facilitate this comparison in terms of heuristic processes, we briefly examine how beliefs about cues and criteria are represented with classical information theory and the basic operations that can be done with these representations.

Belief representation

Beliefs about the value of a cue or criterion are represented as binary values, either present [1] or absent [0]. This entails the assumption that cognitive representations are characterized by discrete states, suggesting that only one set of beliefs or preferences is present at any given point in time. This

¹A realistic analogy would be a prey animal (player A) that can run left or right encountering a predator (player B) that can choose to run left or right.

makes measurement of a state extremely straightforward; a state already exists, so it needs only to be read out of the cognitive system.

Combinations of beliefs about cue or criterion values can then be represented as a series of bits. For example, beliefs about a pair of cues could be represented as [00], [01], [10], or [11]. Our beliefs about a single alternative (e.g. disease present / absent) can be represented using one bit, but if we must make a comparison between two alternatives (e.g. which person is more ill), this also requires two bits. In this case, [01] would indicate that the second person is more ill, [10] would indicate the first person is, and [00] or [11] states would be ambiguous, indicating a non-decision state that would have to be altered in order to directly generate a final choice. To represent beliefs about a pair of cue and a pair of criterion values together, we would need at least 4 bits, yielding 16 possible joint states (e.g. [0000], [0101], or so on).

Note that stringing together bits is also a method for representing more complex cue or criterion beliefs: as the number of bits corresponding to one representation increases, the precision it can provide doubles as well. The greater precision in representations trades off with greater storage and processing demands, as more bits and larger logic gates are required to represent and operate on more precise beliefs.

Much of the work integrating psychology and formal information theory has looked at the issue of storage capacity or discrimination in terms of the number of bits that can be used(Beebe-Center et al., 1955; Garner, 1953; Pollack, 1952), processing chunks of information (Baddeley & Hitch, 1974; Chase & Simon, 1973; Miller, 1956; Wickelgren, 1979). Though not often discussed explicitly in terms of bits, similar discrete-valued representations underlie current slots models of visual working memory as well (Alvarez & Cavanagh, 2004; Awh et al., 2007; Luck & Vogel, 1997). We next visit how chunking and other processes might take place by establishing the operators that are used to transform bits.

Logic gates

The three most basic operations that can be performed on bits or pairs of bits are the NOT (\neg) , OR (\lor) , and AND (\land) transformations, which can be represented either as logical propositions or as circuits (Figure 1). The NOT operator simply inverts one bit, changing [0] (false) \mapsto [1] (true) or [1] \mapsto [0]. The OR and AND gates operate on pairs of bits, returning a [1] if one (for OR) or both (for OR and AND) inputs are [1], and returning [0] otherwise.

By stringing together these three operators, it is possible to specify any classical logic operation. For example, suppose we wanted to set a third bit to one if two initial bits were both one – or more concretely, say hello to someone if you recognize both their face and their gait. This could be



Figure 1. Circuit representation and input-output mapping of basic AND, OR, and NOT gates. Input bits come in the left side and yield outputs out the right side according to the pattern given below each gate.

written as a circuit or as an equation, where response R_1 can be written as a function of cues for facial recognition F, gait G, and initial response intention R_0 .

$$R_1 = (F \land G) \land (R_0 \lor \neg R_0) \tag{2}$$

Note that we include R_0 because a person could initially want to say hello even if they don't wind up recognizing a person's face and gait. In essence, R_0 can build in initial beliefs or response biases. Heuristic models of behavior do not ordinarily incorporate initial beliefs. However, initial beliefs or states have proven particularly important for cognitive process models such as those built using the frameworks of signal detection and random walk models of decisionmaking (Green & Swets, 1966; Link & Heath, 1975; Pleskac & Busemeyer, 2010). In terms of information theory models, the initial state serves as the locus for a person's internal beliefs, which change across consideration of cues and control a person's behavior.

Aside from AND and OR, there exist four other common 2-bit gates: NOR, NAND, XOR, and XNOR. These are often used because NOR gates or NAND gates can by themselves be used to compute any classical logic operation without relying on any other operators (Peirce, 1880; Büning & Lettmann, 1999). However, since we can use AND, OR, and NOT to construct any of these other operators, we leave the rest alone here for simplicity.

When used in tandem, these 2-bit gates can implement any sort of logical rule. However, this has two drawbacks. The first is that a person needs to have many different kinds of gates in order to implement the variety of processing rules they are likely to need. This means that the brain has to produce structures that implement each of the 2-bit gates as well as organize them in sensible ways. While not prohibitively difficult, it perhaps does not provide great credence to systems of 2-bit gates as good descriptions of choice architecture.

Another limitation of the 1- and 2-bit gates we have discussed so far is that they are not reversible, meaning that one cannot infer the inputs to the gates from their outputs because the gates produce fewer bits than they receive. This means that some inputs are are simply lost, letting the metabolic cost of computing them go to waste (in a computer, the energy dissipates as heat, but it is not clear what the biological consequences would be) unless they are stored somewhere in memory.

There is a simply solution to both of these problems, which is to use a 3-bit, reversible gate that can implement any logic operation by itself (or rather with multiple iterations of itself). One such gate is the Toffoli gate, shown in Table 1. This operator checks that two bits are in the state [11] and flips a third bit if so, yielding [110] \mapsto [111] and [111] \mapsto [110] but leaving the triplet unchanged otherwise. The full truth table for this gate is shown below.

	Outputs									
Inputs	000	001	010	011	100	101	110	111		
000	1	0	0	0	0	0	0	0		
001	0	1	0	0	0	0	0	0		
010	0	0	1	0	0	0	0	0		
011	0	0	0	1	0	0	0	0		
100	0	0	0	0	1	0	0	0		
101	0	0	0	0	0	1	0	0		
110	0	0	0	0	0	0	0	1		
111	0	0	0	0	0	0	1	0		
Table 1										

Input-output mapping of a Toffoli gate

The Toffoli gate can actually be strung together with many other Toffoli gates, together allowing them to perform any logical operation and therefore implement any classical processing rule. It is thus referred to as a *universal* gate. This gate is also reversible, meaning it does not lose any information as it is processing incoming bits (though the biological or psychological implications of information loss are not clear). The Toffoli gate could therefore be mass-produced and arranged as needed, making it potentially useful as a computational tool describing how information is processed in the brain. We return to this issue a bit later on, as quantum logic also provides universal, reversible information processing gates like the Toffoli gate.

Relation to heuristics

Heuristics are typically modeled in terms of classical logic, which has made it well-suited to modeling using rulebased cognitive architectures like ACT-R (Marewski et al., 2010; Schooler & Hertwig, 2005), as well as circuit and equation representations (Martignon & Hoffrage, 2002). For our purposes it is useful to examine how heuristics would be implemented as a circuit. Because it can illustrate the basic principles we focus on the recognition heuristic.

The recognition heuristic is a well-studied single-variable decision rule that relies on recognition alone to make a judgment about an unknown criterion value (see for example Goldstein & Gigerenzer, 2002; Marewski et al., 2010, 2011; Oppenheimer, 2003; Pachur & Hertwig, 2006; Pohl, 2006; Volz et al., 2006). When people are given a two-alternative, forced-choice question, the recognition heuristic states: If one of the two objects is recognized and the other is not, then infer the recognized objects has the higher value with respect to the unknown criterion value (Goldstein & Gigerenzer, 2002).

Like many heuristics, recognition exploits information in the environment. In this case, the reason we recognize objects is because of some other mechanism or variable. For example, the cities we read and hear about (and therefore recognize) tend to be the more populous cities (Goldstein & Gigerenzer, 2002; Schooler & Hertwig, 2005). Thus, recognition of a city can be a valid cue of city population. As Goldstein & Gigerenzer (2002) went on to show the recognition heuristic allows people with little knowledge to in some cases outperform for knowledge people in making inferences, yielding the less-is-more effect.

Figure 2 illustrates how the recognition heuristic can be implemented using the 3 basic logic gates (NOT, AND, OR). A string of 4 bits is fed in on the left side, corresponding to the two recognition values of the items $(c_1 \text{ and } c_2)$ and a person's initial beliefs or preferences about their values on the criterion $(b_1 \text{ and } b_2)$. As with the basic logic gates shown in Figure 1, bits are fed into the various logic gates by moving from left to right along the circuit. One bit value is sent to multiple locations where the circuit forks. The output of each gate is fed forward as a bit until reaching the right side of the circuit, where beliefs are measured or evaluated to determine the next response or action.



Figure 2. Example implementation of the recognition heuristic, where b_1 and b_2 correspond to a person's beliefs about the criterion for items 1 and 2 (b'_1 and b'_2 give the revised beliefs), and c_1 and c_2 correspond to whether items 1 and 2 are recognized, respectively.

To be more specific, let us focus on the top half of the diagram and assume that $b_1 = 1$ (e.g. believe that city 1 is large) and $c_1 = 0$ (don't recognize city 1). The [1] for b_1 goes to two places: directly to the lower AND gate, and through a NOT gate (turning into a [0]) to the upper AND gate. The

cue c_1 goes to both of the AND gates as well. In this case, both AND gates will fail – the upper gate receives [00] and the lower one receives [10]. This means that both will pass a [0] along the circuit to the OR gate to their right. Since the OR gate then receives [00] as inputs, it will out put a [0], which is the revised belief $b'_1 = 0$ (city 1 is not large). The bottom circuit works in exactly the same way for a different cue-criterion belief pair (b_2 and c_2) to produce b'_2 .

The construction of this circuit essentially copies the cue values c_1 and c_2 onto beliefs b'_1 and b'_2 , ignoring the initial values b_1 and b_2 . Note that these beliefs still must be mapped onto a decision. There is an additional step that must be taken to transform the final beliefs b'_1 and b'_2 onto actions. If a circuit produced $[b'_1b'_2] = [10]$, it would result in a decision in favor of the first alternative, and producing state [01] would result in a decision favoring the second alternative. On the other hand, states $[b'_1b'_2] = [11]$ or [00] could result in a guess if there were no more cues available, or be fed back into the circuit as $[b_1b_2]$ with another set of cues for the next step.

The result of this circuit is largely unsurprising, as it just implements the recognition heuristic it is supposed to run. However, the diagram is useful for several reasons. One reason is that take-the-best (Gigerenzer & Goldstein, 1996) can be understood as a generalization of the one-cue decision rule shown in Figure 2. Take-the-best is also a well-studied heuristic that ranks cues by their validity and then uses them one by one to discriminate between alternatives (e.g., Bergert & Nosofsky, 2007; Bröder, 2000; Gigerenzer & Goldstein, 1996; Newell & Shanks, 2003; Lee & Cummins, 2004). In take-the-best, each cue processing would pass through another circuit identical to the one shown in Figure 2, only with the next set of cues and revised beliefs as inputs. As long as there were more cues available and no response had been made (checks which could also be done using logic gates), the decision-maker would continue iterating through similar circuits. Thus, a classical circuit representation of take-thebest can be constructed using multiple repetitions of the one shown in Figure 2.

Another useful aspect of the circuit is that it illustrates some important properties of a classical logic approach. First, the output of the circuit is determined as soon as belief and cue values are put into it. While it is possible to modify our framework to assume that the actual logic gates are probabilistic or that people are accessing their cue or criterion beliefs with some noise, the basic behavior of these logic circuits is deterministic.

An additional thing to note is that the gates in the diagram above must be executed in serial. For each belief-cue pair, the OR gates strictly follow the AND gates, and revised beliefs b'_1 and b'_2 must be calculated before the next step can begin. While larger gates can be constructed to address this issue, a person would have to somehow compute the full joint inputoutput relationship of such a gate, a mapping which we have shown grows exponentially with the number of cues.

Finally, the inputs must be determined in order to be transformed using the logic gates. Each gate only accepts binary inputs, meaning that any uncertainty regarding a person's beliefs or the perceived cue values must be resolved before using the circuit. Previous work has assumed that if cue value is unknown, it is treated as if it were zero (Gigerenzer & Goldstein, 1996) or simply skipped altogether (Gigerenzer et al., 1991; Hoffrage et al., 2000). Alternatively, as with the recognition heuristic, there might be some threshold applied to divide continuous values into [0] or [1] based on whether or not it exceeds a critical value (see e.g. Brandstätter et al., 2006; Luan et al., 2011; Pleskac, 2007; Schooler & Hertwig, 2005).

Unfortunately, each of these three characteristics of classical logical processing seems to deviate from the empirical evidence. Experimental evidence has long suggested that decision-making is better described as a stochastic than deterministic process (Audley, 1960; Balakrishnan & Ratcliff, 1996; Busemeyer & Townsend, 1993; Davidson & Marschak, 1959; Simon, 1959), that strategies can involve parallel processing of cues (Cave & Wolfe, 1990; Egeth et al., 1972; Townsend & Wenger, 2004), and uncertainty is used rather than ignored in decisions (Hogarth, 1987; Simon, 1959). The continuous has already caused issues for classical logic-based heuristic and ACT-R theories of recognition, as some nonlinear shapes of typical receiving operator characteristic curves [ROCs] suggest that recognition may not be discrete (Anderson et al., 1998a; Qin et al., 2001; Schooler & Hertwig, 2005; Yonelinas & Parks, 2007) (but see also Malmberg, 2002). Oddly, even theories that posit discrete recognition rely on some underlying continuous value, such as familiarity (see e.g. the theory proposed by Bröder & Schütz, 2009). We are ambivalent on the particulars of this debate, but the recurrence of continuous cue values in models of recognition processes certainly suggests the existence of underlying continuous (or at least graded) representations.

Quantum models of cognition

Although the classical Boolean logic approach provides one basis for constructing cognitive models of cue-based decisions, this is not the only way to approach these models. Generalized Boolean algebras and semi-ordered lattices serve as less restricted mathematical foundations from which to build models (Manes, 1976). Though these can take a number of forms, we focus on a relatively simple and wellstudied alternative based on quantum logic. This approach is particularly useful because many of its formal properties directly address the shortfalls of classical logic that we covered in the last section.

8

KVAM

Rules and violations of classical logic

Recall that Boolean logic is defined by binary-valued representations which are transformed using intersection / AND (\wedge), union / OR (\vee), and negation / NOT (\neg) operations. Three of the axioms of Boolean logic are commutativity, distributivity, and complementarity. Formally, these mean that for events *A*, *B*, and *C*, the following equalities hold:

Commutativity: $B \land A = A \land B$ Distributivity: $A \land (B \lor C) = (A \land B) \lor (A \land C)$ (3) Complementarity: $A \land \neg A = 1$

These three axioms taken together imply that the following relation must also hold:

$$(A \land B) \lor (A \land \neg B) = A \land (B \lor \neg B) = A$$
(4)

These laws have been violated directly or indirectly in empirical studies of human judgment and decision-making. For example, work on order effects in sequential questions has shown violations of commutativity – in these studies, the order of a pair of responses that a participant is asked to make is varied, which leads to different probabilities of responding to the same question based on whether it is asked first or second (Hogarth & Einhorn, 1992; Trueblood & Busemeyer, 2011; Wang & Busemeyer, 2013; Wang et al., 2014).

Work on decision-making under uncertainty has also uncovered violations of distributivity and complementarity. If a person prefers one item over another (preference state C) under one condition B, giving $C \wedge B$, as well as under the complementary condition $\neg B$, giving $C \land \neg B$, then she should prefer the same item when she is unsure of the state of the world, $C \wedge (B \vee \neg B)$. Though it is formulated based on the distributivity and complementarity axioms of classical logic, this is also referred to as the sure thing principle (Savage, 1954). This principle was violated in a number of empirical studies - notably, participants who were offered two sequential gambles were willing to take the second gamble if they knew they had won the first gamble or if they knew they had lost the first, but chose not to take the second gamble if they were unsure of the outcome of the first one (Tversky & Shafir, 1992a). Similarly, in the two-player Prisoner's Dilemma game, participants chose to defect if they knew the other player had chosen to defect or if they knew the other player had chosen to cooperate, but did not choose to defect if they were unsure whether the other play had cooperated or defected (Busemeyer et al., 2006; Tversky & Shafir, 1992b). This is referred to as a *disjunction effect*, which violates Equation 4 above.

Empirical violations of the law of total probability have also appeared on a number of tasks, including inferential decisions like those that are featured in studies of fast and frugal heuristics. For instance, Townsend et al. (2000) showed participants a series of faces. On some of the trials, they had to categorize the faces as friendly or hostile (i.e. decide on a cue value) and then decide whether to be friendly or defensive. On other trials, they simply decided whether to be friendly or defensive. Oddly, participants tended to be more defensive when they made no categorization than when they did (collapsed across 'friendly' and 'hostile' categorization decisions).

Classical models, including heuristic models, typically assume that participants should reach the same decision (act friendly or defensive) on the two different types of trials. This is because the logical rule shown in Equation 4 implies that the marginal probability of being defensive should be the same across the two different conditions. Being more defensive on trials without a categorization than with categorization is a violation of a classical logic rule. This empirical violation has, however, been replicated several times and shown to be consistent with a quantum model of the decision process (Busemeyer et al., 2009; Wang & Busemeyer, 2016b).

Order effects and implications for modeling frameworks. The axioms we outline above are implicitly assumed in models of cognition that are based on classical logic information processing and classical probability. This encompasses both classical logic-based heuristics as well as more complex frameworks like Bayesian models of belief updating. Neither one would predict order effects a priori. For example, asking a person what they believe the value of a cue is (e.g. "Does city X have a major league baseball team?") should not affect their beliefs with regard to a criterion ("Is the population of city X greater than 500,000?") according to classical models, but it is well-established that these and similar effects occur across tasks and contexts (Atmanspacher & Römer, 2012; Busemeyer et al., 2006; Hogarth & Einhorn, 1992; Kerstholt & Jackson, 1998; Khrennikov & Haven, 2009; Kvam et al., 2013, 2015; Moore, 2002; Pothos & Busemeyer, 2009; Townsend et al., 2000; Trueblood & Busemeyer, 2011; Tversky & Shafir, 1992b; Wang et al., 2014; Wang & Busemeyer, 2016b).

Interestingly, work by (Holyoak & Simon, 1999) has suggested that beliefs about cues can be impacted by beliefs about choice criteria (e.g. the verdict on a person's guilt) as well as by the values of earlier cues (Glöckner, 2007). Such interactions between cues and criteria provide a substantial stumbling block for models of cue-based decisions, and particularly for Bayesian belief updating models that suggest that people reason unidirectionally from cues to criteria. Quantum logic, as we show later, proposes a mechanism by which this could occur based on entangled cues and beliefs.

These sorts of order effects and other violations of classical probability rules are often taken for granted partly because adding additional components to the models can sometimes account for violations, so individual violations are seen as unique psychological phenomena rather than a systematic issue with our approach to modeling cognition. However, such modifications must often be made post hoc and tend not to be sufficiently general to explain even a single type of violation across all of the tasks in which it appears.

A reasonable question might be raised as to why we introduce a quantum framework rather than shifting to a classical probability or Bayesian approach to describing cognition (Griffiths et al., 2010). While such approaches can certainly add stochastic elements, parallel processing, and continuous beliefs, they still adhere to the rules of classical probability. As such, the violations we described above remain extremely problematic. Furthermore, shifting to a Bayesian framework prevents us from implementing simple rules using logic gates. Instead, it requires a more complex processing infrastructure to deal with joint probability distributions over all combinations of beliefs, updating of internal probabilities (or probability mass / density), as well as consideration of factors such as cue covariances. For these reasons, we do not visit Bayesian models or other classical probability models of decision behavior in much detail here.

This leaves us with the question of whether we should continue to ground heuristics and other models in classical logic and probability frameworks, modifying them as needed to account for new violations of classical laws. Alternatively, it may be useful to consider a framework whose first principles are well-suited to describing and predicting behavior that violates these rules. In the following sections, we pursue the latter option by proposing quantum logic as an alternative framework for constructing information processing models.

Quantum logic

One solution to account for the persistent violations of classical laws is to add additional cognitive assumptions to the models, such as belief revision (Hogarth & Einhorn, 1992) or motivated reasoning (Tversky & Shafir, 1992b). We pursue an alternative solution to the persistent violations of classical laws, which is to apply quantum probability to modeling cognition and ask if these models based on their first principles better account for decisions. Models built using this quantum framework have been used to model violations of commutativity (Trueblood & Busemeyer, 2011; Wang & Busemeyer, 2013), distributivity / the sure thing principle (Pothos & Busemeyer, 2009), and the law of total probability (Busemeyer et al., 2009; Khrennikov & Haven, 2009; Kvam et al., 2015). Quantum models are able to violate these axioms because they use different formal representations, transformations, and measurement operators to describe a system.

The formal underpinnings of quantum probability models can be used to construct a quantum system of logic. While the application of quantum logic to model cognitive processes is a recent innovation, these sorts of models have been used previously in quantum computing and information, so the basic principles are well-established (see Nielsen & Chuang, 2010; Barenco et al., 1995; Busemeyer & Bruza, 2012). Instead of readable bits and logic gates, quantum logic is based on qubits, measurement operators, and unitary processing gates.

Here, we introduce these basic concepts, show how they can be used to re-construct heuristics and other rule-based models, and examine what diverging predictions they make when we compare them to heuristics based on classical logic. Our coverage of quantum probability models is not exhaustive, but here we aim to give readers enough information so that they have a basic understanding of how and why quantum logic can be used to model information processing, and hopefully provide sufficient background that readers could construct several heuristics on their own (for a more complete introduction and tutorial, see Busemeyer & Bruza, 2012; Yearsley & Busemeyer, 2015).

States & measurements. Rather than using bits, quantum logic represents pieces of information as qubits. The fundamental difference between qubits and bits is that qubits can exist in a continuous superposition state of both 0 and 1 simultaneously rather than discrete values. This superposition state can be represented as a vector in two dimensions, where each dimension corresponds to a value the qubit can take - for example, one dimension could correspond to a "cue present" response and the other to a "cue absent" response. The two vectors $|0\rangle$ and $|1\rangle$ serve as bases for describing the superposition state, so a qubit can be represented as a linear sum of the two, $B = b_0 |0\rangle + b_1 |1\rangle^2$ The coefficients b_0 and b_1 represent probability amplitudes, which are used to determine the probabilities of obtaining a $|0\rangle$ or a $|1\rangle$ when the qubit is measured (i.e., when a decision about the cues or criteria is made). Note that these probability amplitudes can actually take on both real and imaginary or complex values, but for simplicity all of our examples use strictly real numbers.

The superposed qubits also correspond to some of our intuitions about how we represent states. Rarely do we have a definite belief about whether a particular city has greater or fewer than 1 million inhabitants, but it often seems the case that we have a feel or 'fuzzy' impression of the city being large or small (somewhat similar to fuzzy set theory Klir & Yuan, 1995). The continuous superposition state provides a formal representation that describes these sorts of vague beliefs as well as makes a number of empirically testable predictions.

In order to observe an output from a quantum state, as in making a binary decisions, one needs to measure it. That is, in quantum models of cognition a judgment or a decision is

²This is referred to as "ket" notation. For simplicity, one can assume (for example) that $|0\rangle$ is the x dimension and $|1\rangle$ is the y dimension as shown in Figure 3.

a measurement applied to a person's cognitive state (Busemeyer & Bruza, 2012). Measurement of a quantum state is taken by projecting it onto the corresponding basis (or bases), with the probability of obtaining a particular value given by the squared length of the current state along that basis vector (see Figure 3). The value of a single qubit is measured by projecting it onto the $|0\rangle$ or $|1\rangle$ basis, with the probability of obtaining 0 or 1 given by the squared length of the qubit along the $|0\rangle$ or $|1\rangle$ basis vector, respectively. Note that by virtue of being unit length, the probabilities of obtaining a [0] or a [1] sum to unity, $|b_0|^2 + |b_1|^2 = 1$. Once a qubit is measured, it 'collapses' on the observed state, reducing the probability amplitude in the complementary state to zero. In essence, measurements create a definite state from an indefinite one.

For example, an initial state may begin in a superposition, $\psi = b_0|0\rangle + b_1|1\rangle$. If a measurement is applied and obtains the outcome |1\rangle, the new state must be revised to reflect this. As a result, the state after measurement ψ' is no longer in a superposition state but rather equal to the definite state, $\psi' = |1\rangle$.



Figure 3. Representation and measurement of a qubit $|B\rangle$. While b_0 and b_1 can be negative or even complex, their squared lengths must sum to unity.

Critically, quantum logic models based on qubits are inherently stochastic, but they can be restricted so that they can only be in definite states $|0\rangle$ or $|1\rangle$. For example, we could set $b_1 = 0$ or $b_1 = 1$ in Figure 3 and obtain fully deterministic behavior from the qubit. In this case, they will behave exactly like classical deterministic bits. This makes the qubit representation a general case of the bit. As we show later, information processing in the quantum framework is also a general case of classical logic implementations – they can do everything exactly as classical deterministic heuristics does or they can be permitted to vary more freely to provide closer,

stochastic descriptions of human behavior.

As we will see in the next section, a person's beliefs about cue or criterion values need not be resolved before being used, allowing indefinite (uncertain) beliefs or cues to be used in decision strategies rather than ignored or assumed away. The quantum notion of uncertainty is a bit different than that of the classical models – rather than uncertainty regarding external events like population size due to a lack of knowledge, a person can have internal uncertainty about their own beliefs. We visit the distinction between types of uncertainty in more detail in the discussion.

As a result of the use of superposition states, we obtain both stochastic behavior as well as incorporate uncertainty regarding cues and criterion values. Therefore, the introduction of the qubit addresses two of the major deficiencies that were elucidated when we examined the classical information theory approach to heuristics.

Compound states. In the classical bit-based framework, we could simply string together sequences of bits in order to form bit strings. The same can be done in the quantum framework – a series of coordinates can describe a string of qubits – but by using qubits to represent a person's beliefs about cues and criteria, it is actually possible for them to interact with one another. In such a case, not only do cues affect criterion judgments, but revision or measurement of criterion beliefs can affect a person's beliefs about the cues as well (similar to what is described in hindsight biases or parallel constraint satisfaction; see Hoffrage et al., 2000; Glöckner, 2007).

Multiple qubits can be combined by assigning a probability amplitude to each of the combinations of their potential values. For example, a pair of qubits can be described using four basis vectors, describing the 4 combinations of measurements one could obtain: $|00\rangle$, $|01\rangle$, $|10\rangle$, and $|11\rangle$. The superposition is a linear combination of these four possibilities, $\psi = c_{00}|00\rangle + c_{01}|01\rangle + c_{10}|10\rangle + c_{11}|11\rangle$. When the fist qubit is measured, the probability of obtaining a 1 for the first qubit (for example) is $|c_{01}|^2 + |c_{11}|^2$. After measuring a 1 for the first qubit, the qubit would collapse on states $|01\rangle$ and $|11\rangle$.

The simplest way to combine (\otimes) two superposition states (either individual qubits or sets of qubits) is to take their outer product, $\psi_a \otimes \psi_b$, yielding a superposition over the possible combinations of their states. ³ If we combine qubits A and B using this method, we will obtain a joint state $\psi_{ab} = a_0b_0|00\rangle + a_0b_1|01\rangle + a_1b_0|10\rangle + a_1b_1|11\rangle$. As with individual qubits, the sum of the squared lengths of the coefficients will sum to one if they are combined in this way.

Critically, a superposition over two qubits can exist without combining them via the outer product, such that ψ_{ab} =

³The Kronecker product of two matrices is a generalization of the outer product from vectors to matrices. The Kronecker product of two matrices X and Y can be easily computed in Matlab with the function kron(X, Y) or in R with kronecker(X, Y, ...).

 $c_{00}|00\rangle + c_{01}|01\rangle + c_{10}|10\rangle + c_{11}|11\rangle$, where $c_{00} \neq a_1b_1$ and/or $c_{01} \neq a_1b_2$, and so on. The two qubits are then said to be *entangled*. In this case, measuring one qubit affects the measurement of the other. As a result, an entangled state cannot be separated into two independent qubits.

For example, imagine trying to measure the entangled state ψ_{12} when $c_{00} = \sqrt{.5}$, $c_{11} = \sqrt{.5}$, and $c_{01} = c_{10} = 0$. If we measure the first qubit and obtain a 1, this forces the joint (entangled) state to collapse on $|10\rangle$ and $|11\rangle$. The new state would be $\psi'_{12} = |11\rangle$ (since there is no amplitude on state $|10\rangle$). Note that this affects the potential measurements on the second qubit! We can no longer obtain a value of 0 when we measure the second qubit because we obtained a particular measurement on the first qubit. This has important consequences for how cues are processed and become paired with beliefs. Ultimately, it leads us to predict both order effects and provides a structural explanation of why the hindsight bias phenomenon occurs.

Conceptually, a superposition state representation of our beliefs about cues or criterion values corresponds to a type of uncertainty about their value. With a superposition representation of an unknown binary cue value, a person considers both possibilities simultaneously rather than directly assigning discrete 0/1 values or probabilities to the outcomes. This is in stark contrast to the assumption made in heuristics, where unknown values are assumed to be $0.^4$ In essence, a qubit-based representation allows people to utilize uncertainty in their decisions rather than forcing them to assume it away or abandon cues. Additionally, it suggests that new information interacts with uncertain states differently than individual, certain states. The ultimate effect of this difference in representation becomes more clear as we see how qubits interact with the gates used to process information.

U-gates. Recall that in binary logic, we used logic gates to operate on bits of information. In the quantum logic framework, we instead use unitary gates [U-gates] to transform qubits or sets of qubits. A unitary gate can be described by the operations it performs (rules it implements) or in the form of a matrix operator that is multiplied by its input qubit(s). Several of the important unitary gates are shown in Figure 4.

Conceptually, applying a unitary operator to a qubit belief state is similar to rotating it in a multidimensional space – in fact, rotations of vectors in real space are unitary transformations. This means that they are fairly flexible in terms of how they can change a qubit. We therefore focus on a few key gates that are most useful for modeling heuristics, but there are more that can be used and adapted based on the strategies a person could be implementing.

Of course, the flexibility of these U-gates is not always a benefit. In many cases, as with the U_{θ} gate shown in Figure 4, they introduce free parameters that must be set both by the decision-maker as well as by the modeler attempting to esti-

mate a U-gate-based model. This makes them computationally more complex than the classical gates and deterministic models, which carries the potential risk of overfitting.

One of the simplest U-gates that can be used is called the Pauli-X gate. This gate simply flips a qubit, changing its value along $|1\rangle$ to its value along $|0\rangle$ and vice versa. For example, a qubit $\psi = [\sqrt{(.9)}; \sqrt{(.1)}]$ would become $\psi' = [\sqrt{(.1)}; \sqrt{(.9)}]$. The matrix representation of this gate is

Pauli-X:
$$\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$
 (5)

The Pauli-X gate can be interpreted as a NOT gate, as it inverts whatever qubit is put into it.⁵

As in the classical framework, the quantum NOT gate can be applied in series with other gates, allowing for serial processing of series of bits. However, it can also be combined with other operations in order to perform more complex transformations of multiple qubits. For example, we use a *controlled-NOT* or CNOT gate heavily in our implementation of the recognition heuristic and take-the-best. This operator is particularly useful because it can implement rules that are structured in an "if-then" way. It applies a NOT (qubitflip) transformation to turn on a qubit that is off $(|0\rangle \rightarrow |1\rangle)$ if a particular cue is observed, making it ideal for modeling the cue-triggered rules specified by heuristics.

The CNOT gate is shown in Figure 4 along with two other important gates that we use to implement quantum logic heuristics. This gate checks the value of a first qubit (top two rows of a 2-qubit [4-row] state), and inverts the entries of a second qubit (bottom two rows of a 2-qubit state) if it has a value along |1).

Another important gate is called the *Hadamard* gate, shown on the right side of Figure 4. This gate takes a single qubit ($\psi = |0\rangle$ or $\psi = |1\rangle$) and rotates it by 45 degrees. If the qubit is in a definite state – as it would be after being measured – this gate randomizes its next measurement so that the probabilities of obtaining a 1 and obtaining a 0 are equal. Application of this gate allows a person to get random measurements, in essence allowing them to choose randomly between a pair of alternatives. It can therefore be applied to the final belief states when all cues have been exhausted in order to permit random guessing behavior. Accordingly, this gate is applied at the final step of take-the-best (shown below) as a method of generating guesses when no more cues are available.

⁴This assumption is not made explicitly, but the outcomes of unknown "?" cue values is practically equivalent to values of zero in terms of choice outcomes.

⁵On top of serving as a NOT gate, this is one of the Pauli spin matrices (Pauli-X) that serve as the basis of many unitary transformations, along with Pauli-Y and Pauli-Z transformations (Nielsen & Chuang, 2010). To apply it, one multiplies the gate's matrix by the qubit vector, $\psi' = \sigma_x \psi$.

KVAM





Illustration:





	$\cos(\theta)$	$-\sin(\theta)$		
U _e =	$sin(\theta)$	$\cos(\theta)$		

Action: Rotates qubit B_1 by an angle θ .

Application: Can be used to adjust beliefs partially, especially if controlled by cues that are not entirely convincing.



Circuit:



Matrix form:

	[1	0	0	0
–	0	1	0	0
О _{см} –	0	0	0	1
	0	0	1	0

Action: Flips qubit B_1 if qubit C_1 has value along $|1\rangle$.

Application: When fed a joint qubit $(C_1 \otimes B_1)$, uncertainty in cue is transformed into uncertainty in belief. (see Appendix A)

 $B_1 \rightarrow H \rightarrow B_1'$

Hadamard gate

Illustration:



Matrix form:



Action: Rotates a definite-state qubit B_1 by 45 degrees.

Application:

Used to generate a random response (guess) when no more useful information is available.

Figure 4. Representations and applications of several important U-gates.

More on U-gates. A more general version of the Hadamard gate, the (Euler) rotation operator U_{θ} , is shown on the left side of Figure 4. Unlike the gates we have discussed so far, this gate utilizes a free parameter θ that controls how far the gate rotates a qubit. While doing so adds complexity to the model, it may in some cases be justified. For example, this gate allows for qubits to be partially rotated depending on how convincing a cue is to a particular individual.

As with the CNOT gate, the operation of any unitary transformation can be controlled by a separate qubit. For example, one could control the operation of the rotation matrix by forming the larger operator $U_{C\theta}$:

$$U_{C\theta} = \begin{bmatrix} 1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & \cos(\theta) & -\sin(\theta)\\ 0 & 0 & \sin(\theta) & \cos(\theta) \end{bmatrix} = \begin{bmatrix} I_2 & \mathbf{0}\\ \mathbf{0} & U_{\theta} \end{bmatrix}$$
(6)

Other controlled operations can be implemented by substituting different unitary transformations for U_{θ} in the bottom right corner.

In addition, transformations can be controlled by multiple qubits using doubly-controlled gates. For example, one might note that the Toffoli truth table from Table 1, when put in matrix form, is also a unitary gate. It is in fact a CCNOT gate, where a third bit is flipped if the first two have values along $|1\rangle$.

Interestingly, the CCNOT gate is a special case of the

Deutsch gate (Deutsch, 1989), and can be used in conjunction with definite-state qubits to implement any operation that exists in classical logic. The quantum logic framework we have described can therefore implement any classical logic heuristic, but its more general structure allows for it to be potentially more empirically accurate.

It is important to note at this point that qubits can implement bit-based logic (by being set to the definite states $|0\rangle$ or $|1\rangle$) and U-gates can implement any classical logic operation. This permits quantum logic to execute any decision strategy that can be created using classical logic. In many cases, it may be useful to use the classical logic implementations – for example, when one desires computational simplicity and fewer free parameters – but this does not speak against a quantum logic approach more generally. Instead, the multitude of cases where quantum logic can provide a better description of behavior suggest that it can be used as the general case, while deterministic or classical heuristics can serve as simpler special cases.

In addition to controlled gates, multiple qubits can be processed in parallel by simply combining the gates together and combining the qubits together. In the same way that qubits can be combined by taking the Kronecker product of the two vectors, two unitary transformations can be combined by taking their Kronecker product as well, $U_{12} = U_1 \otimes U_2$. The new unitary matrix can then be applied to a (Kroneckercombined) pair of qubits, with each unitary transformation affecting one of the qubits individually.

$$U_{12} \cdot \psi_{12} = (U_1 \otimes U_2) \cdot (\psi_1 \otimes \psi_2)$$

= $(U_1 \cdot \psi_1) \otimes (U_2 \cdot \psi_2)$ (7)

The joint qubit state will have each of its constituent qubits transformed by the corresponding gate: U_1 will affect only ψ_1 , and U_2 will affect only ψ_2 . This provides a straightforward method for processing information in parallel, avoiding one of the other critical limitations of serial-order heuristics.

Even more complex gates can be formed either by specifying them manually or by taking Kronecker products of other gates, including specific transformations corresponding to learned rules or template gates that implement more common logic operations. This could serve as a basis for spontaneous rule combination that is believed to occur as part of learning in cognitive production systems (see e.g. Anderson, 1982).

More complex gates allow us to approach more complex and task-specific problems as well. One issue with models using only a few qubits is that they cannot map onto a substantial number of responses, meaning that the simple gates we have examined so far cannot produce more continuous responses like confidence. However, larger gates could implement cue-controlled unitary operations such as random walks, allowing for transformations that unfold over time and over a large number or continuous states. Later, we also review how this approach could provide a means for modeling other processes like evidence accumulation and heuristics like tallying and weighted additive rules.

One final note of interest on unitary operators is that they all meet the condition $U^{\dagger}U = UU^{\dagger} = I$ where *I* is the identity operation and U^{\dagger} is the complex conjugate transpose of *U*. This makes unitary transformations inherently reversible – any operator can be undone by simply applying its complex conjugate transpose to its output. Note, however, that prior states cannot be reverse-computed using this approach once they are measured, as measurement alters the cognitive state from the one obtained immediately after cue processing. The particular implications of reversibility for cue-based decisions are not yet clear, but reversible gates do provide the basis for constructing production systems that can accomplish backward induction tasks (Anderson, 1982; Anderson et al., 1984; Tarrataca & Wichert, 2012).

Uncertainty and entanglement. So far we have examined the effects of definite state qubits on the transformations, but what if one or more of the qubits is in an indefinite or uncertain state? First, consider the effect of uncertain cues on rule-based processing. Suppose that we have one cue that indicates that a disease is present if true, and not present if false. This cue serves as the first qubit, and our beliefs about the presence of the disease are indicated by the second qubit B. Before we have any information, we are in state $|0\rangle$, believing that the disease is not present. Let us also assume that we have some uncertainty regarding the cue value C – this could occur if the evidence is inconclusive, if the cue is inaccessible, if we have insufficient data to fully diagnose the cue value, or if we want to represent it as a continuous value. In this case, we represent C as a superposition of $|0\rangle$ and $|1\rangle$, $C = c_0 |0\rangle + c_0 |1\rangle = [c_0; c_1]$, with $c_0, c_1 \neq 0$.

In order to apply our unitary operator, we compute the initial state across qubits, $CB = c_0|00\rangle + c_1|10\rangle$. Note that $B = |0\rangle$, reflecting our initial belief that the disease is not present. Because the rule is an if-then rule, we use a CNOT gate U_{CN} so that if some control condition is satisfied (presence of cue) then a unitary operator is applied to some target (belief states are changed). More formally, the initial state *CB* is transformed yielding a revised state, $CB' = U_{CN}(CB)$. The revised state is $CB' = c_0|00\rangle + c_1|11\rangle$. Note that when we dissect this state, the probability of measurements on the cue have not changed, $Pr(C = 0) = |c_0|^2$ and $Pr(C = 1) = |c_1|^2$. But the probability of measuring our beliefs about the criterion has changed to match the probability of the cue being measured, so that $Pr(B' = 0) = |c_0|^2$ and $Pr(B' = 1) = |c_1|^2$.

In effect, the uncertainty about the cue has been transformed into uncertainty about whether the disease is present. This means that the probability of making a particular decision or response is directly affected by our representation of uncertainty in the information we are using. Furthermore, not only are measurements of our belief about the criterion now stochastic, but the cue and our beliefs are now entangled, so that observing the perceived value of the cue will affect our beliefs and observing our beliefs will affect the perceived value of the cue. For example, if we were to measure the cue by projecting the state onto $|10\rangle + |11\rangle$ (with probability $|c_1|^2$), a subsequent measurement of our beliefs would be guaranteed to result in a positive answer because the two qubits cannot be in states $|01\rangle$ or $|00\rangle$. However, if we were to measure our beliefs first, there would be a probability $|c_1|^2 < 1$ that we would obtain a positive measurement of "disease present."

This has a particularly interesting consequence: once a cue belief and criterion belief are entangled, reducing uncertainty about the criterion value also reduces uncertainty about the cue value. This means that feedback about a criterion directly affects cue beliefs. This is precisely the mechanism that is assumed to generate the hindsight bias in heuristic models of the hindsight bias (Hertwig et al., 1997; Hoffrage et al., 2000). We revisit this point later when discussing empirical evidence for quantum heuristics.

Application to heuristics

So far, we have illustrated some of the important elements of the quantum framework. These are used to compose quantum models of all sorts, but we are most interested in quantum logic constructions of fast and frugal heuristics. Here, we present quantum versions of the recognition heuristic and the take-the-best heuristic as well as discuss how cuecriterion belief entanglement can lead to a hindsight bias. The formal structure of the recognition heuristic and takethe-best models can be found in Appendix A and Appendix B, respectively.

In the most simple case, quantum logic can simply be used to reconstruct the classical Boolean structure of existing heuristics. In this case, one would set all qubits for beliefs and cue values to one of the definite states, $|1\rangle$ or $|0\rangle$. For instance, one could obtain the same choice behavior from our implementation of take-the-best as other classical implementations of take-the-best by by eliminating uncertainty about cues and beliefs by setting all cues and beliefs to $|0\rangle$ or $|1\rangle$.

One consequence of this is that quantum logic heuristics must by definition be able to perform at least as well as classical logic heuristics – the former can mimic behavior of the latter. However, by using uncertainty and different sets of processing rules (U-gates), quantum logic heuristics should be able to outperform the accuracy of classical logic ones. In this way, the performance of typical classical logic heuristics serves as a sort of lower bound on the maximum potential of quantum logic heuristics. We provide an example of how incorporating uncertainty into the structure of the heuristic can improve accuracy at the end of the next section on the recognition heuristic.

It is important to note that the improved performance conferred by quantum logic heuristics does not necessarily make them better descriptive models of behavior, which is the primary goal of our integration. However, selection pressures on both our evolved and learned strategies should push them toward more successful ones, suggesting that the higherperformance quantum logic heuristics would likely be used if they can be developed. Furthermore, if a person outperforms a classical logic heuristic in terms of their choice accuracy, this does not necessarily imply that heuristics are unable to account for that person's behavior. Instead, the additional accuracy conferred by modeling the person's choice behavior using a quantum logic heuristic (as opposed to a classical one) could explain the gap in performance in such cases, allowing the heuristic to be more descriptively accurate by virtue of its improved performance.

Recognition heuristic

We can reconstruct the recognition heuristic discussed in previous sections using quantum logic gates and qubits. A more formal description of the model with the involved mathematics is contained in Appendix A; here we provide a simpler walk-through of how the quantum logic recognition heuristic functions. Recall that the recognition heuristic strategy is used for binary choice, where decision-makers check if they recognize a pair of alternatives, and choose one alternative if it is recognized and the other is not.

Suppose a person is deciding which of two cities has a larger population, Bakersfield (CA) or Atlanta (GA), based on which of the two cities they recognize. Using the quantum logic framework, we can represent a person's initial beliefs about the size of the cities as qubits B_1 for Atlanta and B_2 for Bakersfield. Then we can set their recognition cue values of each city as qubits C_1 for recognizing Atlanta and C_2 for recognizing Bakersfield.

We can model a person's initial beliefs to be unbiased regarding which city is larger by setting B_1 and B_2 to $|0\rangle$ (i.e. believe neither city is larger before assessing recognition). Of course, recognition is likely to be quickly computed and used in this problem, but before doing so a person has no information on the cities and it is therefore reasonable for them to have no biases to respond one way or another.

Next, we must set some values for the recognition cues. In principle, these values would be informed via theories of recognition memory using perhaps ACT-R (Anderson et al., 1998b), MINERVA-DM (Dougherty et al., 1999), or some other theory to predict a familiarity signal. However, for our purposes, we leave this back-end specification open. Suppose that a person definitely recognizes Atlanta, so recognition is in definite state $|1\rangle$, $C_1 = [0; 1]$. By contrast, suppose they are uncertain about whether they have seen Bakersfield somewhere before, so recognition is low at $C_2 = [\sqrt{(.8)}; \sqrt{(.2)}]$. If they are asked if they recognize Bakersfield, there is a 20% chance of a positive response and 80% chance of a negative one (weaker or stricter recognition criteria could adjust the bases used to evaluate the statement "I recognize Bakersfield" and systematically affect these probabilities).

This recognition information needs to be processed in order to change the person's beliefs. To do so, we pair the cue with criterion beliefs (C_1 with B_1 and C_2 with B_2) to obtain two qubit pairs, C_1B_1 and C_2B_2 . These joint qubit states describe a person's state in terms of 4 possible combinations, recognize / not recognize × believe larger / smaller. These joint qubits are paired with one another by taking the Kronecker product of the two constituent qubits.

Each of these joint qubit states needs to be transformed so that a person's uncertainty about recognition informs their uncertainty about the criterion. To do so, the cues are then used to update the person's beliefs using the controlled Ugate U_{CN} that we described above. This transformation takes the uncertainty in the cue values C_1 and C_2 and transfers it onto the beliefs B'_1 and B'_2 . The revised joint qubit states $C_1B'_1$ and $C_2B'_2$ are computed by multiplying the unitary matrix by the combined cue-belief states.

$$C_1 B_1' = U_{CN}(C_1 B_1) \tag{8}$$

$$C_2 B_2' = U_{CN}(C_2 B_2) \tag{9}$$

The resulting states $C_1B'_1$ and $C_2B'_2$ each describe a superposition over beliefs about both the cue and criterion. For example, $C_1B'_2$ will be $[\sqrt{(.8)}; 0; 0; \sqrt{(.2)}]$, yielding a 20% probability of responding that Bakersfield is recognized and large in population ($|11\rangle$), and a 80% chance of responding that Bakersfield is not recognized and small in population ($|00\rangle$). Interestingly, according to this model, there is no chance of responding that the city is recognized but small or not recognized but large based purely on recognition as a cue.

Once a person's final beliefs are computed, they can make their final decision between Atlanta and Bakersfield. To do so, we need to evaluate the person's beliefs about the city sizes, B'_1 and B'_2 . If B'_1 is evaluated as $|1\rangle$ (100% chance with our example values) and B'_2 is evaluated as $|0\rangle$ (80% chance in our example), the decision maker will select Atlanta as the larger city. Conversely, if B'_1 is evaluated as $|0\rangle$ (0% chance) and B'_2 is evaluated as $|1\rangle$ (20% chance), Bakersfield is selected as the larger city. In all other cases, recognition will be insufficient to decide between the two because the decision maker's beliefs about the two cities are measured to be the same.

If a decision-maker still needs to decide, they may choose randomly between the cities. In this case, they could invoke the Hadamard gate shown in Figure 4. To do so, they would simply feed one qubit in (e.g. B''_1 , the measured state obtained from B'_1) and measure the revised state that comes out

of the Hadamard gate. If the revised state is measured as $|1\rangle$, they choose Atlanta (50% chance), and if it is measured as $|0\rangle$, they choose Bakersfield. In our example, the overall choice proportions are 90% choosing Atlanta (80% chance of choosing it based on recognition + 20% × 50% chance of choosing it randomly) and 10% choosing Bakersfield (20% × 50% chance of choosing it randomly).

Note that the final decision can be made by computing a joint state from the decision maker's final beliefs, $B'_1B'_2$, and measuring both beliefs simultaneously. This joint state represents their combined beliefs about the two cities, and contains all of the information needed to make a decision. For example, they choose Atlanta if the joint state is $|10\rangle$, Bakersfield if the joint state is $|01\rangle$, and randomly (by feeding their beliefs through a Hadamard gate) if the joint state is $|00\rangle$ or $|11\rangle$.

Adding the use of uncertain cues also allows for potentially greater performance on this task as well. For example, we could make a reasonable assumption that larger cities have a tendency to have uncertainty leaning in the direction of positive recognition, whereas smaller cities may have a tendency to lean toward negative recognition. Because both beliefs are uncertain, a classical recognition heuristic will force them to bypass this cue and choose randomly between cities. However, if uncertainty can vary along a spectrum and the recognition qubit value is correlated with the criterion value, even uncertain recognition can be useful. For example, suppose a person is uncertain about whether they recognize both Atlanta and Bakersfield in our example above: Atlanta recognition qubit [$\sqrt{.9}$; $\sqrt{.1}$] is compared against Bakersfield qubit $[\sqrt{.8}; \sqrt{.2}]$. If the quantum recognition strategy is used, it will result in a correct decision $(18\%|01) + 50\% \cdot 72\%|11) + 50\% \cdot 2\%|00\rangle = 55\%$ of the time, as opposed to the 50% one would obtain by bypassing these uncertain cues and simply guessing.

Heuristics can therefore reap substantial benefits simply by incorporating uncertainty into their structure. However, the influence of different cues can also be moderated by using different logic transformations – for example, less useful cues could apply U_{θ} transformations that rotate beliefs toward the criterion value only partially. This would allow the decision strategy to incorporate the validity or reliability of different cues into their choice probabilities, guarding against inferences made using poor information.

Take-the-best heuristic

Just as with the classical implementation of take-the-best, the quantum implementation of take-the-best can be seen as a generalization of the implementation of the recognition heuristic. To introduce the model, imagine again that a person is trying to decide whether Atlanta (A) or Bakersfield (B) has a larger population. For simplicity, let us assume that they recognize both cities and can recall some information about them, so instead of relying on recognition they are using a take-the-best strategy to prioritize and examine cues (though of course recognition could be incorporated formally as a prior step or integrated as simply another cue in a heuristic). A diagram of an implementation of take-the-best with 3 cues in the hierarchy is shown in Figure 5 (excluding the recognition cue as an initial step).

The three cues that the person uses are the presence of a league baseball team in the city, the presence of a university in the city, and whether there is a subway system in the city. Note that the cues are still selected and ordered on the basis of validity (though information or success is potentially a better alternative, see Hilbig, 2010; Newell et al., 2004), so the quantum logic implementation is subject to all of the same benefits and drawbacks of stratifying cues in this way.

We can set the initial beliefs so that prior to considering any information, a person does not believe either city to be larger in population, $B_1 = B_2 = [1; 0]$. Suppose that they are unsure but likely to respond yes if asked whether Bakersfield has a baseball team, $C_{1,1} = [\sqrt{.2}; \sqrt{.8}]$ (in reality, Bakersfield has a minor league team) but they know that Atlanta has a major league team, $C_{2,1} = [0; 1]$. When these cue values are paired with their respective beliefs ($C_{1,1}$ with B_1 and $C_{2,1}$ with B_2) and passed through the CNOT gate U_{CN}^2 , the uncertainty in the cue values is transformed into uncertainty in beliefs, so that $B'_1 = [\sqrt{.2}; \sqrt{.8}]$ and $B'_2 = [0; 1]$.

At this point, a person's beliefs are measured using the measurement projector(s) **M**. The measurements can result in one alternative being favored over the other ($|01\rangle$ or $|10\rangle$) or a person's beliefs can be evaluated as the same for both alternatives (a tie, $|00\rangle$ or $|11\rangle$). Even if the unresolved beliefs about the two alternatives (superposition states) are not the same, there is still the possibility that they will be equal when measured. In our example, there is a 20% chance that they will say that they believe Atlanta is larger (and that Bakersfield does not have a professional baseball team), a 0% chance that they do not resolve the decision at this stage as they believe both cities are large because they have baseball teams.

After the measurements, the belief states are projected back onto $|0\rangle$ and the person continues onto the next cues. Suppose that they know that both cities have a university – processing cues $C_{1,2}$ and $C_{2,2}$ (both equal to $|1\rangle$) will not result in any decision when they are used to change the person's beliefs.

In contrast to the other cues, the decision-maker's beliefs about the presence or absence of a subway system in the two cities may be largely uncertain but doubtful, so that they are unlikely to say that Atlanta has a subway system $(C_{2,3} = [\sqrt{.8}; \sqrt{.2}])$ and slightly more likely to say that Bakersfield has a subway system $(C_{1,3} = [\sqrt{.7}; \sqrt{.3}])$. When the uncertainty from these cues is transformed into uncertainty in beliefs about the sizes of the city, a person will have a 14% chance of selecting Atlanta as having a larger population (20% chance of saying Atlanta has a subway times 70% chance of saying Bakersfield does not), a 24% chance of selecting Bakersfield as having a larger population (30% chance of saying Bakersfield has a subway and Atlanta does not), and a 62% chance of selecting neither (20% times 30% chance of saying both have a subway plus 70% times 80% chance of saying neither has a subway).

Finally, a person who has still not chosen an alternative proceeds to the Hadamard gate, which allows them to randomly choose between Atlanta and Bakersfield (50% chance each). Final choice proportions can be calculated by adding the probabilities of choosing Atlanta or Bakersfield across the different steps of the heuristic. In our example, they will choose Atlanta as the larger city 61.5% of the time and Bakersfield as the larger city 38.5% of the time. The formulas for calculating the choice proportions are provided on the bottom-right of Figure 5). Of course, this is merely an example. These choice proportions would change depending the cues used (e.g. if recognition was not 100%) and the degree of uncertainty associated with each one.

The formal construction of the logic gates and cue-belief combinations is provided in Appendix B. We have also made Matlab code available that carries out the computations for this example, available on the Open Science Framework at osf.io/yph97. It is useful to note again that the quantum implementation of the take-the-best heuristic seamlessly integrates stochastic behavior into the rule, demonstrating that the deterministic nature of previous implementations of the take-the-best are not necessarily a property of the heuristic itself, but of the information processing theory in which the heuristic is grounded. It is also useful to highlight that past implementations of take-the-best or similar rules have struggled with how to deal with uncertain or unknown cue values. Sometimes these unknown cue values were treated as missing (see Figure 3 in Gigerenzer et al., 1991) or treated as equivalent to a non-occurrence (see Figure 3 in Gigerenzer & Goldstein, 1996). Instead of needing these ad hoc assumptions, as we have shown, quantum information theory handles uncertainty in cue values with its first principles.

Parallel take-the-best and compensatory strategies

Although the serial structure of heuristics is often an advantage (as some information can be ignored), it can at times be inefficient. For example, when high-validity cues are infrequently available, they still must be inspected in order to get to diagnostic cues. In a related vein, it has been frequently found that gaining substantial expertise on a particular task tends to move strategies from a cue by cue method of processing information to a parallel or holistic processing pattern (Ackerman, 1988; Dreyfus, 2004). As such, it should be possible within our information processing framework to



Figure 5. Simplified circuit representation of a 3-cue, 2-alternative quantum logic version of take-the-best. Mathematical construction of cues and belief representation, information processing, and measurement matrices as well as a more complete mathematical form of this diagram are presented in in the Appendix.

move from a serial to a parallel method of processing cues (similar to how it is described in ACT; Anderson, 1982).

As we saw with cues and beliefs and the compound U_{CN}^2 gates, it is straightforward in the quantum framework to combine cues and processing gates together into a single state and unitary transformation. This allows us to construct parallel versions of take-the-best as well as implement compensatory heuristics like linear or additive strategies (Hogarth & Karelaia, 2007). An example of a parallel implementation of a heuristic is shown in Figure 6. Instead of feeding in cue and belief qubits in sets of 4, all 12 are fed into the processing gate U_{CN}^6 at the same time. This allows for all cues to affect beliefs simultaneously in parallel.



Figure 6. Simplified circuit representation of a parallel quantum logic heuristic. Cue $(C_{i,j})$ and criterion $(B_{i,j})$ beliefs are fed into the information processing gate U_{CN}^6 and then measured with the operator(s) M, yielding choice frequencies x and y.)

However, this has the drawback that the measurement operator \mathbf{M} becomes more complex. In fact, different mea-

surement operators can implement different heuristics. For example, we can implement serial measurements identical to those presented in the serial implementation of take-thebest, measuring first beliefs affected by the first cue ($B_{1,1}$ and $B_{2,1}$) followed by the second and third. This would execute take-the-best but save time in terms of serial cue processing, though measurement would require the same operations as before.

Alternatively, we could measure all beliefs at the same time. For example, one could measure all cases where more beliefs favor option A than option B (e.g. $B'_{1,1}B'_{2,1}B'_{1,2}B'_{2,2}B'_{1,3}B'_{2,3} = |101011\rangle$, $|101010\rangle$, etc.) or more favor option B than option A (e.g. $B'_{1,1}B'_{2,1}B'_{1,2}B'_{2,2}B'_{1,3}B'_{2,3} = |010111\rangle$, $|010101\rangle$, etc.) – this would implement the tallying heuristic. One may want to include a Hadamard gate at the end so a person would choose randomly in cases where measured beliefs in the number of cues favoring A and B are tied.

Other compensatory heuristics could be implemented in a similar way. There are a large number of potential strategies that could be created by adjusting the processing or measurement operators, so we do not go into too many here. Linear weighted additive heuristics could substitute $U_{C\theta}$ for U_{CN} , and adjust θ for the various decision weights. Or one could adjust the measurement operator to check for dominance conditions (where 1+ beliefs favor option A and none favor B) before using other rules. Either one of these could be implemented in the same parallel way, and rules could be easily adjusted by modifications to the processing and measurement matrices.

Similarly, different measurement operators could examine combinations of cues simultaneously – for example, it could project onto qubits describing beliefs influenced by cues 1 and 3. This allows for compound beliefs or cues to be used in the decision process, as proposed by Garcia-Retamero et al. (2007).

Just as individual cues can control partial rotations, they

KVAM

can also control other unitary operations. For example, an individual cue could control a random walk operation, changing how beliefs are distributed across multiple levels of perceived evidence. This would allow for quantum logic models to predict cue-based confidence judgments similar to the the confidence models of Kvam et al. (2015) or Wang & Busemeyer (2016a).

Quantum logic heuristic predictions

Across all of the heuristics we have examined, there are a number of running themes that arise. Because of the structure of quantum logic that we use to implement heuristics, we obtain uncertainty in beliefs, partial transformations resulting in probabilistic decisions, uncertainty in cue values, parallel cue processing, and cue-belief entanglement. These reflect the 5 new predictions of quantum logic heuristics:

- Decision-making (measurement) can be made stochastic due to the superposition representation of beliefs.
 For a given set of cues and initial beliefs, unlike in the classical framework, a single set response is not always guaranteed.
- Cues need not be entirely convincing. Depending on the validity or subjective credibility of a cue, it can apply a partial transformation, shifting a person's beliefs toward but not fully into the criterion state it indicates.
- Cue values can be indeterminate, either by being inaccessible, incomplete, or continuous to a decisionmaker. This uncertainty is used to form beliefs rather than ignored or assumed away.
- Unitary gates are combined to operate on multiple qubits in parallel using a single transformation, suggesting that switching from serial to parallel processing of cues in decision strategies should be a rapid transition. In addition, parallel cue processing allows for more flexible rules to be applied via different measurement operators, including both dominance-based conditions and cue combinations.
- Once a cue is processed, it may be come entangled with a person's beliefs. This results in order effects, where evaluating a cue value affects subsequent beliefs about a criterion value, and evaluating beliefs can reduce uncertainty about a cue value. Therefore, responses after processing an uncertain cue depend on the order of subsequent measurements.

Discussion

By integrating the quantum logic construction of cues, beliefs, and information processing with simple heuristic rules for utilizing information, we gain a number of theoretical and empirical advantages over existing approaches describing human behavior in cue-based decisions. In addition, it allows us to make a number of new, testable predictions that provide avenues for future research. In the following sections, we review some of these insights, outline points of divergence between classical and quantum logic approaches to heuristics, and suggest directions for further investigations.

Benefits of heuristics for quantum logic

The benefits of using simple heuristics are by now wellestablished (see e.g. Gigerenzer & Todd, 1999; Hertwig et al., 2013; Todd et al., 2012), and a quantum logic formulation of them preserves many of these advantages. This allows quantum logic heuristics to inherit many of the present benefits of fast and frugal heuristics. Though not an entirely exhaustive list, we review some of the most important ones here.

Structure and rules. Quantum logic as a framework is indifferent to how particular strategies are put together – alone, it does not offer an explanation of behavior. It simply provides a set of representations and computations, building blocks with which we can construct strategies. Therefore, quantum logic relies on established heuristic strategies to guide construction and provide context and theoretical power to its implementation.

As we hope to have shown, the building blocks of heuristics – information search, stopping rules, and a decision rule (Gigerenzer, 2004) – can be mirrored almost exactly in the quantum logic constructions of heuristics. For example, the serial version of quantum take-the-best (Figure 5) gathers information cue-by-cue beginning with the highest validity, stops when a measurement obtains a discriminating decision state, and results in selection of the alternative that state favors. These three rules are still essential components of constructing a quantum logic heuristic, but must navigate the slightly different cue and belief representations and information processing circuits.

There is no default structure to measurement operators either, and as we saw in the parallel heuristic implementation, different operators can result in the implementation of many potential decision strategies. Interestingly, the deliberate use of measurement operators offers a new way to explicitly model the decision rule. Because the cognitive system is always in a definite state in the classical approach, the decision rule is often an after-thought. In the framework of quantum logic, decision making is actually a constructive process where a cognitive state is created when a person's beliefs are measured, so the particular measurement (the basis in which the state is created) is particularly important. Of course, quantum logic does not provide the heuristic strategies by itself, only the means to implement and adjust the heuristics. **Simplicity.** One of the greatest strengths of heuristics is that they can deliberately ignore information in order to obtain faster, more accurate, or more frugal outcomes. This is passed on to the quantum logic implementations of these same heuristics, though of course parallel cue processing strategies sometimes may use unnecessary information as with tallying or other compensatory heuristics.

This also allows quantum logic heuristics to avoid the calculations necessary to compute the covariance matrix between all of the cues that are used. Instead, the heuristics typically exploit statistical structures of the environment such as valid cues like recognition (Goldstein & Gigerenzer, 2002), so-called non-compensatory cue structures (Gigerenzer, 2004; Martignon & Hoffrage, 2002), or via dominance and cumulative dominance (Simsek & Buckmann, 2015). These properties allow the heuristics to avoid the costly process of setting and re-adjusting decision weights based on all of the co-occurrences of available cues. And as the research on fast and frugal heuristics shown, these restricted strategies sometimes perform just as well or even outperform these decision processes that do (Gigerenzer & Todd, 1999; Todd et al., 2012).

Finally, deterministic heuristics actually provide a particular special case of their quantum logic implementation – namely, one where only definite states are used and logic gates are restricted to deterministic forms. In many cases, this may make them more parsimonious descriptions of behavior by minimizing the number of free parameters that have to be estimated in the model. Assuming particular forms of representation and processing also simplifies the task of the decision-maker, who does not have to worry about their internal levels of uncertainty or about the strength of transformations they apply to their beliefs (e.g. deciding on different values of θ for the U_{θ} gate).

Ecologically adaptive. As we showed in previous sections, quantum logic can implement any strategy that is done in classical logic. Therefore, we know that quantum logic heuristics can achieve performance that is at least equal to that of classical logic heuristics. As such, they inherit a fairly high degree of base performance by implementing the recognition heuristic, take-the-best, or other well-documented strategies (Gigerenzer et al., 1991, 1999; Hogarth & Karelaia, 2007; Pachur et al., 2011).

However, quantum logic versions of these heuristics could potentially achieve even greater performance in some environments. For example, if the uncertainty in the cue beliefs reflects some true (or at least seemingly-probabilistic) unpredictability of the environment, participants will actually match the probabilities of their occurrence in their behavior. Matching probabilities in the environment is actually optimal in many environments, such as choosing a foraging patch (Kennedy & Gray, 1993; Krebs, 1978), and this optimal behavior would be generated by the heuristic structure combined with its quantum logic implementation.

As we suggested before, quantum logic heuristics' improved performance may confer benefits in terms of the descriptive accuracy of heuristic strategies. If people's behaviors adapt to the uncertainty present in their environment, then modeling heuristics using quantum logic components may better reflect the improved strategies we develop by incorporating uncertainty. It also brings instances where people out-perform classical heuristics back into consideration as heuristic strategies, as the addition of quantum logic components can potentially bridge the gap between the limits of classical logic heuristics and observed choice behavior.

Benefits of quantum logic for heuristics

The benefits of modeling cognitive processing using heuristic accounts and the benefits to decision-makers of using heuristics are already quite well-established (see for example Gigerenzer et al., 1999). Here, we focus more on how merging quantum logic with heuristics makes them differ from classical logic heuristics and what benefits doing so confers.

Uncertain and continuous cues and beliefs. Even if cues are certain, a person may apply a partial rotation like U_{θ} to change their beliefs when a cue is not entirely convincing. In some cases, this will result in a person not reaching a decision after considering a definite cue. So even if a cue is diagnostic, a participant using a heuristic may not stop after inspecting it and instead continue to consider other information. In order to examine this, one would have to check trials on which a cue that should not have been used (based on the cue hierarchy in some non-compensatory decision strategy) and see if this cue was unexpectedly correlated with the final decision a participant made on those trials. This is essentially the approach that was used to test the priority heuristic (see e.g. Fiedler, 2010) where later cues (i.e., the difference in worst-case probabilities or the best outcome in this case) were shown to affect choice when according to the classical interpretation of the priority heuristic it should not have. Oddly, this behavior of quantum logic heuristics suggests that instances of apparently compensatory behavior may arise during the use of non-compensatory heuristics if cue values are partially uncertain.

There is already evidence for a continuous underlying structure to cues. For example, we reviewed recognition, where setting a threshold on some continuous recognition value in order to make it binary yields linear ROC functions (e.g. 'high-threshold' models Malmberg, 2002). Empirical data that suggest nonlinear ROCs would necessitate a continuous underlying recognition value (Yonelinas & Parks, 2007), but many alternative proposals for linear ROCs (e.g., threshold models) also rely on a continuous underlying cue, such as memory strength or familiarity (Bröder & Schütz, 2009). Such a representation is well-captured by continuous

qubit values, much like the continuous values used in signal detection theory.⁶ Which cues are continuous and to what extent the continuous nature of cues is used when available is an open question.

Parallel cue processing and combination. As we described in previous sections, multiple cues and beliefs or multiple processing gates can be easily combined with one another in the quantum logic heuristics. They can then be processed altogether, and different decision rules can be applied to the person's revised beliefs. This can potentially result in a much faster decision process, as inspection of each cue is not dependent on the diagnosticity of a cue higher in the inspection hierarchy. In principle, quantum logic allows for rapid rule combination and parallelization of sets of logic gates.

Unfortunately, gathering evidence that cues are processed in this way may be difficult. Serial and parallel processes can be notoriously hard to discriminate, depending on the structure of the proposed models (Townsend, 1971, 1990). However, quantum logic would suggest that serial gates can be created by combining them via a Kronecker product. This leads to the empirical prediction that the transition from a serial to a parallel procedure could be quite rapid rather than occurring over gradual periods of time.

Although little work has been done to our knowledge on how heuristic processes may change over time and with repetition, some support for this prediction can be found in expertise literature. In several studies in this area, participants seem to transition from periods of serial, declarative rules when they are only familiar with the task to parallel, procedural rules when they have become very good at it (Ackerman, 1988; Dreyfus, 2004; Beilock & Carr, 2001). The periods of transition are typically characterized by rapid increases in processing speed or accuracy as well as spontaneous rule combination. These rapid shifts are pervasive enough that they have been used to demarcate different stages of expertise acquisition (Taatgen, 2005; Anderson, 1982), hinting that rapid gate construction may be a reliable component of learning to perform a task.

The parallel processing of cues also means that decision rules can be applied by using different measurement operators after all information processing has completed. This opens up new strategies, as a person could process all cues at once and then check dominance conditions or use multiple cues in combination with one another to trigger a choice one way or the other (Garcia-Retamero et al., 2007).

Stochasticity. The switch from a strictly deterministic to a stochastic system of cues and beliefs is perhaps the most controversial feature of a quantum logic structure for heuristics. In addition to making the system a bit more complex, it also means that strategies must be graded in a different manner. Rather than the rate of adherence to a strategy as a marker of empirical accuracy, we must shift to a likelihood-

based method of model (strategy) evaluation. We note reasons for this in Appendix C.

Although either change constitutes a substantial shift, they are both in line with modern models of the decision-making process. There is strong evidence that some stochastic elements are not only desirable but necessary in order to produce the choice proportions and distributions of response times that we observe in empirical data (Townsend & Ashby, 1983) (see also Busemeyer & Bruza, 2012; Busemeyer & Townsend, 1993; Hogarth, 1987; Simon, 1959).

However, it is also fair to say that models with deterministic elements can provide more parsimonious accounts in many cases. This is commensurate with fixing parameters to reduce flexibility – in a Bayesian sense, it would be equivalent to restricting the priors on parameters of a (probabilistic) model in order to improve its predictive accuracy in a particular environment. Sticking to a simpler deterministic model is particularly desirable when there is little choice data for estimating free parameters.

Many models of the decision-making process have been built with adding stochasticity in mind, deliberately building in elements that make decisions probabilistic. Several authors have incorporated probabilistic behavior into heuristics, either as part of an evidence accumulation process (Lee & Cummins, 2004), part of the evaluation process (Rieskamp, 2008), or as random selection of different heuristics (Rieskamp & Otto, 2006; Davis-Stober & Brown, 2011). However, these are post-hoc modifications to the theory, which in a sense indicates that there may be something wrong with purely deterministic models of cognition. To be fair, the many potential sources of noise in behavior make it difficult to tell if a particular underlying cognitive system is behaving in a deterministic or stochastic way. It may well be that cognitive processes are inherently deterministic and that we have to build in some functions that produce error. Without being able to identify a consistent source of the largely probabilistic behavior that arises in cue-based decisions, however, modeling variation in outcome as a function of superposition is arguably more parsimonious and no less likely than most other accounts.

Interestingly, evidence suggests that there are substantial advantages to behaving unpredictably in a stochastic way. For example, in game theoretic scenarios it is often beneficial to adopt a mixed strategy, performing different behaviors with some amount of randomness (Colman, 2003; Smith, 1974, 1982). Taken a step further, using explicitly quantum strategies can improve a player's expected payoffs in a variety of environments, even invading evolutionarily stable classical strategies (Eisert & Wilkens, 2000; Iqbal & Toor, 2001; Meyer, 1999). Similar to mixed strategies, it may be adaptive to scale the probability of different responses based on inputs

⁶In fact, quantum signal detection theory (Baldo, 2013) mimics its behavior and would produce essentially the same results.

or predicted outcomes, leading to behaviors like probability matching (Bliss et al., 1995; Shanks et al., 2002; Thompson, 1933). Substantiating the predicted random behavior, there appear to be mechanisms in the brain that generate probabilistic behavior or choices (Tervo et al., 2014) much like a Hadamard gate would implement.

Theory integration and conclusions

Each of the theories we have integrated here provides some benefit to the other, but perhaps more interesting is that putting them together raises new questions and new perspectives on existing ideas. Two important ones are the predicted effects of entanglement of cue and criterion beliefs, and the incorporation of uncertainty into the structure of the cue processing models.

Entanglement. The integration of quantum logic and heuristics allows for the new possibility that cue and criterion beliefs can become entangled during information processing. Specifically, it suggests that if an indeterminate cue influences beliefs (via a U-gate operation), evaluation of the cue should affect subsequent evaluation of beliefs and information about the criterion should affect beliefs about the cues. This results in two interesting effects: order effects of belief measurements regarding cues and criteria, and effects of criterion information on cue beliefs. The latter may underlie the hindsight bias phenomenon.

To give a more concrete example of an order effect regarding cue and criterion beliefs, consider the situation a person is gauging whether to attend a movie based on whether they recognize it. The decision-maker may be unsure of whether they recognize a movie (to different extents), and so processing this cue may make them unsure of whether they want to attend the movie. However, if we force them to respond that they do or do not recognize the movie, this should influence the probability that they decide to see it or not. As a result, if we contrast responses under two conditions - one where we ask if they recognize it and then if they want to see it, and another where we only ask if they want to see it - we should obtain violations of the law of total probability (see Equation 4). Results of similar sequential-choice experiments suggest that this may be the case (Kvam et al., 2015; Townsend et al., 2000; Wang & Busemeyer, 2013; Wang et al., 2014; Wang & Busemeyer, 2016b).

Cue-belief entanglement also results in an effect of criterion information on cue beliefs. Getting information about the criterion values results in a transformation or collapse of an entangled cue-criterion state, changing a person's beliefs about the cue as well. This results in modified beliefs about the value of the cues compared to when a person initially processed the cue values. As a result, when a person revisits their (revised) cue beliefs to calculate their confidence in the criterion, this value may be greater than when they processed the cue initially. As a result, retrospective confidence following criterion information can be higher than prospective confidence before criterion information, creating a hindsight bias.

This explanation of hindsight aligns closely to previous heuristic descriptions of the hindsight bias, such as the Reconstruction After Feedback with Take-the-best [RAFT] model (Hertwig et al., 1997; Hoffrage et al., 2000). This model of the hindsight bias posits that after a decision is made, feedback about the criterion value (reducing uncertainty about criterion beliefs) leads to inferences about cue values that were potentially uncertain at the time of a decision (reducing uncertainty about cue beliefs). Formally, a quantum logic model of take-the-best or the recognition heuristic would say that collapsing the cue-criterion belief state on a definite state of criterion beliefs increases a person's belief in the cue. This is precisely the effect of entanglement – changing criterion beliefs should affect cue beliefs once they have been entangled together through a U-gate.

Retrospective confidence, in turn, would be determined by the revised cue beliefs, leading to higher confidence after receiving feedback - a hindsight bias. Therefore, a hindsight bias falls readily out of the quantum logic heuristics we described in the previous section. Although we have not discussed it in detail here, there are multiple ways to model confidence in the quantum logic framework. We could do so by using the revised cue beliefs to directly yield confidence (as done in Trueblood & Busemeyer, 2011) or to trigger a separate confidence-generating operation (as in the random walk model proposed by Kvam et al., 2015). We should emphasize that this prediction of a hindsight bias is not the result of adding auxiliary assumptions to the process model like cognitive reconstruction or motivated self-presentation to the model (Hawkins & Hastie, 1990). Instead, the prediction arises from the first principles of the quantum logic model of heuristics developed here, which raises interesting questions as to the hindsight bias' status as a bias at all.

Uncertainty and entropy

Markov and quantum models make a distinction between mixed and superposition states. Mixed states give probabilities of observing different results (decisions, confidence judgments), but these probabilities are predicated on the uncertainty of the observer or modeler. On the other hand, superposition states represent measurement uncertainty, which makes both the decision-maker and any observers uncertain about a person's actual state. This distinction is particularly interesting because it illustrates how different types of uncertainty correspond to distinctions made in work on heuristics between types uncertainty.

In a classical framework, we might discriminate between *epistemic uncertainty* – an inability to predict future outcomes primarily due to a lack of knowledge – and *aleatory uncertainty* – an inherent inability to predict future outcomes

due to their unpredictable nature. For example, epistemic uncertainty might describe our uncertainty about the official census population of a city in terms of our lack of knowledge, while aleatory uncertainty might describe our essentially irreducible uncertainty about the exact number of people within the city's boundaries at a particular time.

The mixed states described in classical probability principally correspond to reducible epistemic uncertainty, where a series of bits is only unpredictable if we lack knowledge about whether each entry is a 0 or a 1. Instead, we might only be able to supply some probability that an entry will be 0 or 1. The closer this probability is to 0 or 1, the more predictable the string and its behavior (e.g. how it interacts with logic gates) will be. Formally, the amount of unpredictability of a classical state is referred to as the Shannon entropy (Shannon & Weaver, 1949), and it can be quantified by a number of different metrics. We reduce the entropy of a classical system by knowing more about the state, to the point where we can completely predict the properties and behavior of a bit string if we know what each bit is (0 entropy). Conceptually, then, Shannon entropy embodies the idea of epistemic uncertainty, where our inability to predict the outcomes of events or states can be reduced by knowledge of the system in question (Ascough et al., 2008; Fontana & Gerrard, 2004).

Aleatory uncertainty supposedly arises from inherently stochastic properties of the environment. In a truly deterministic world, we could in theory predict anything given enough knowledge of the current state of the world. In this sense, aleatory uncertainty may not truly exist in a classical deterministic system. However, integrating heuristics with the quantum framework introduces a parallel form of uncertainty called *ontic uncertainty* – uncertainty regarding the true existence of a particular state – which arises from the stochastic interactions between states and measurements of the world. In this framework, ontic uncertainty serves as the principally irreducible type of uncertainty or unpredictability, and therefore may be the true source of what we more commonly refer to as aleatory uncertainty.

Formally, a quantum superposition state such as a qubit possesses unpredictability in the form of von Neumann entropy (von Neumann, 1955). In this case, we can know everything about a state – for example, $\psi = [\sqrt{.5}, \sqrt{.5}]$ – but the state is still entropic in that we cannot perfectly predict the outcome of a measurement. Uncertainty about a qubit string cannot be completely reduced unless every qubit has collapsed on $|0\rangle$ or $|1\rangle$, in which case the von Neumann entropy is 0. Therefore, there can be uncertainty about what state a qubit is in as well as uncertainty about what measurements we will obtain even if we know the state (Bengtsson & Zyczkowski, 2006). Von Neumann entropy therefore provides a 'true' source of unpredictability (ontic uncertainty) that may underlie what we label aleatory uncertainty (Ascough et al., 2008; Fontana & Gerrard, 2004). While it may seem inconvenient for a person in a von Neumann-entropic state to be unable to predict their own behavior (assuming they know their own state), this is advantageous in that it is a source of uncertainty that cannot be reduced by other people either. As we mentioned earlier, this is particularly desirable because it prevents an agent from being exploited when opponents have a great deal of information about the agent's strategies (Colman, 2003; Smith, 1974, 1982).

Conclusions

In this paper, we have examined in detail how quantum logic and heuristics might inform one another. Heuristics provide rules and structure for constructing simple strategies to make decisions based on cues, and quantum logic provides a set of representations and computations that lend empirical accuracy and flexibility to the strategies that are composed.

As a result, both quantum framework and heuristic rules are enhanced by integrating them into quantum logic heuristic models. But perhaps more importantly, putting them together in this way raises new questions. It opens up the possibility of cue-criterion belief entanglement, provides connections between uncertainty in judgment and decision-making and entropy in formal information theory, posits mechanisms for random choice and rule combination, and provides a potential bridge between high-level cognitive theory and lowerlevel neural implementation.

Examining and replacing the formal structures that underlie cue processing models can generate insights that might otherwise be taken for granted. We hope this effort on integrating multiple approaches – covering heuristics, information theory, and classical and quantum logic – has raised new questions, illustrated the importance of critically examining the assumptions that we make when modeling, and offered promising alternative directions for exploring models of information processing more generally.

Appendix A: Formal implementation of quantum logic recognition heuristic

We suppose that beliefs about a criterion (e.g. which of two cities is larger?) are represented by 2 qubits B_1 and B_2 , and the recognition cue values are also represented by qubits C_1 and C_2 . Therefore, the initial state *CB* of the system can be described as an outer product of these four qubits:

$$CB = (C_1 \otimes B_1) \otimes (C_2 \otimes B_2) \tag{10}$$

If we want to flip a criterion belief qubit if a cue is present (city is recognized), then we use the CNOT gate U_{CN} shown in Figure 4 on each of the qubits individually, combining the gate to do so in parallel.

$$U_{CN}^2 = U_{CN} \otimes U_{CN} \tag{11}$$

Then, we apply this gate to the initial 4-qubit state that we obtained by combining the cue belief qubits and criterion belief qubits.

$$CB' = U_{CN,2}(CB) \tag{12}$$

In order to measure the revised beliefs, we must project onto the corresponding belief states. There are 3 projectors to construct: one for "choose Bakersfield," one for "choose Atlanta," and another for "no choice / check another cue." Note that "choose city 1" projector P_1 checks for instances where the second and fourth qubits are evaluated as 1 and 0 respectively, "choose city 2" projector P_2 covers instances where the second and fourth are evaluated as 0 and 1, and the remaining projector P_N covers all other measurement outcomes.

Choose city 1:
$$P_1 = |0100\rangle + |0110\rangle + |1100\rangle + |1110\rangle$$

Choose city 2: $P_2 = |0001\rangle + |0011\rangle + |1001\rangle + |1011\rangle$
Choose neither: $P_N = |0000\rangle + |0101\rangle + |1000\rangle + |0010\rangle$
 $+ |1101\rangle + |0101\rangle + |0111\rangle + |1111\rangle$

The probability of choosing city 1 is $||P_1|C_1B'_1C_2B'_2\rangle||^2$, city 2 is $||P_2|C_1B'_1C_2B'_2\rangle||^2$, and continuing without a response or choosing randomly is $||P_N|C_1B'_1C_2B'_2\rangle||^2$.

In the case that a person does not collapse on a left or right response, the Hadamard gate H from Figure 4 can be applied to any or all of the collapsed qubits in parallel.

This gate maps pure $|0\rangle$ or $|1\rangle$ states to an equal superposition of the two. When a state is measured after applying the gate, it will randomly collapse with equal probability on one of these basis states. In effect, it takes a definite set of beliefs and maps them onto random responses, providing a mechanism for generating random choices that is not present in the classical framework. The Hadamard gate can be applied in a parallel manner using a Kronecker product as in Equation 11.

Appendix B: Formal implementation of quantum logic take-the-best

The diagram shown in Figure 5 is a largely simplified version of how a quantum logic take-the-best is implemented. Given initial beiefs B_1 and B_2 and cues $C_{1,1}$ and $C_{2,1}$, the initial state that is fed into the first gate (U_{CN}^2) is a specific combination of the beliefs and cues, which we refer to as CB_1 .

$$CB_1 = (C_{1,1} \otimes B_1) \otimes (C_{2,1} \otimes B_2)$$
 (13)

Because it is being applied to a pair of cue-belief states, gate U_{CN}^2 is actually a pair of U_{CN} gates combined via Kronecker product, $U_{CN}^2 = U_{CN} \otimes U_{CN}$. Note that the Kronecker

product is distinct from matrix multiplication – U_{CN}^2 is not simply the square of U_{CN} .

In order to process the cues, we simply multiply the unitary matrix by the combined belief and cue state to obtain the revised state $CB'_1 = U^2_{CN}CB_1$.

The measurement **M** consists of 3 projectors: one to calculate the probability of choosing Atlanta $[M_A]$, one to calculate the probability of choosing Bakersfield $[M_B]$, and one to calculate the probability of choosing neither and instead continuing to the next gate $[M_N]$. The projector M_A measures the probability amplitude along dimensions where a person believes that city A is large and B is not (i.e. where $B_1 = 1$ and $B_2 = 0$). It therefore consists of a 16 × 16 matrix of zeroes with ones along the diagonal in rows 5, 7, 13, and 15. Conversely, the projector M_B measures the probability amplitude along states where $B_1 = 0$ and $B_2 = 1$, and consists of a 16 × 16 matrix of zeroes with ones along the diagonal in rows 2, 4, 10, and 12. Finally, M_N projects onto states where $B_1 = B_2$ and consists of a 16 × 16 matrix of zeroes with ones along the diagonal in rows 1, 3, 6, 8, 9, 11, 14, and 16.

The probability of each response is calculated by squaring the length of the projections given by each of the measurement operators.

$$Pr(\text{choose A}) = x_1 = |M_A CB'_1|^2$$
 (14)

$$Pr(\text{choose B}) = y_1 = |M_B C B'_1|^2$$
 (15)

$$Pr(\text{choose neither}) = z_1 = |M_N CB'_1|^2$$
(16)

People who choose A or B are finished, but those who do not choose either have their belief states B_1 and B_2 projected onto $|0\rangle$ and normalized to length 1 (so the probabilities of all of the possible measurements again sum to 1), then continue onto the next step of the heuristic. Their beliefs are paired with a new set of cues to compute a combined state, they process the cues using the same information processing matrix, and the measurement operators are applied to the resulting state.

Combined state $CB_2 = (C_{1,2} \otimes B_1) \otimes (C_{2,2} \otimes B_2)$ (17)

Revised state
$$CB'_2 = U^2_{CN}CB_2$$
 (18)

$$Pr(\text{choose A}) = x_2 = |M_A C B'_2|^2$$
 (19)

$$Pr(\text{choose B}) = y_2 = |M_A CB'_2|^2$$
 (20)

$$Pr(\text{choose neither}) = z_2 = |M_N CB'_2|^2 \qquad (21)$$

This process is repeated until a choice alternative is selected or until a person runs out of cues. In the case that they run out of cues, a person chooses randomly. The random state is generated using a Hadamard gate H:

$$H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1\\ 1 & -1 \end{bmatrix}$$
(22)

Multiplying H by any definite state $B_1 = |0\rangle$ or $B_1 = |1\rangle$ will result in state $B_1 = [\sqrt{.5}; \sqrt{.5}]$. This state is measured by projecting it onto $|0\rangle$ (choose B) or $|1\rangle$ (choose A) – the probabilities of each response will be exactly 0.5.

Parallel heuristic

The U_{CN}^6 gate shown in Figure 6 is formed by taking the Kronecker product of 6 U_{CN} gates.

$$U_{CN}^{6} = U_{CN} \otimes U_{CN} \otimes U_{CN} \otimes U_{CN} \otimes U_{CN} \otimes U_{CN}$$
(23)

The cue-belief state that is fed into it is formed by first combining the criterion belief $B_{i,j}$ with the corresponding cue $C_{i,j}$. Then, all of the cue-criterion belief pairs are combined together to form the overall cue-criterion belief state *CB*.

$$CB = (B_{1,1} \otimes C_{1,1}) \otimes (B_{2,1} \otimes C_{2,1}) \otimes (B_{1,2} \otimes C_{1,2}) \otimes \dots (24)$$

As before, the revised state is computed by multiplying the unitary operator by the cue-criterion belief state, $CB' = U_{CN}^6(CB)$. This revised state is then measured using some set of measurement operators *M* defined by the decision rule specified in a heuristic.

In essence, the U_{CN}^6 gate processes 6 cues at once – 3 corresponding to each alternative – and uses them to modify each of 6 belief states about the criterion. This allows it to process information from many sources all at the same time, but the resulting beliefs (B'_n) have to be measured in non-trivial ways. The measurement operator therefore becomes a particularly interesting point of variation in strategies, as it suggests that the differences in behavior result not from differences in information search, but from the decision rule applied once all cues have been gathered and processed.

Appendix C: Justification for likelihood over adherence-based model evaluation

As far as methods of model evaluation, adherence rates provide a generally poor measure of model performance compared to likelihood-based methods, as it can often overrepresent the true empirical accuracy of a model (Hilbig, 2010). The primary issue with adherence rates is that using this metric presupposes a deterministic model of the world. By using adherence, model performance is graded on its capacity to predict each data point individually rather that its ability to describe the data-generating process. This winds up essentially penalizing all stochastic models, and also runs into problems when applied to grading mixtures of strategies across or within participants.

For example, suppose we want to describe a machine that produces 70% blue and 30% red objects. A deterministic model that always predicts blue objects would have a 70% adherence rate, while a model that predicts 70% blue objects would have a [70% * 70% (correctly predict blue) + 30% * 30% (correctly predict red) =] 58% adherence rate if it made probabilistic predictions or 70% adherence rate if it made its maximum likelihood prediction on every new object produced. However, we would never say that the deterministic model is more descriptively accurate than the model that represents the true stochastic structure of the machine.

Adherence rates have an additional downside when it comes to mixture data as well. For example, it can run into trouble when different participants are using different strategies and the group-level data is analyzed. Suppose that cue 1 is more valid than cue 2, but Participant 1 prefers to use only cue 1 and Participant 2 prefers to use only cue 2 to make their decisions. If cue 1 was available to Participant 1 but not Participant 2, their group-level choice behavior would correspond perfectly to take-the-best even though neither participant was using this strategy. Conversely, if participant 2 While this is a simple example, it illustrates the conceptual challenges of using adherence rates when trying to grade models in the presence of a mixture of strategies. In essence, the best-fitting model to a set of aggregate data may So not only do adherence rates falsely penalize stochastic models, they cannot really be used except to predict homogeneousstrategy data either.

On the other hand, maximum likelihood methods have been both highly successful and widely used as methods of parameter fitting and model comparison in statistics and psychology (Myung, 2003). Potentially more appropriate methods such as Bayesian estimation also incorporate likelihood functions, even if they are not simply maximizing them.

Returning to our 70% blue, 30& red machine and models, we would say that the likelihood of the 70% blue, 30% red machine is $L = Pr(blue)^{(\#blue)} \cdot Pr(red)^{(\#red)}$. The maximum likelihood model, supposing that there are 700000 blue objects and 300000 red objects produced, would be the model positing 70% blue and 30% red objects, while the deterministic model would have probability zero (or closer to zero, if we assume some error that is not equal to 30%). This is a much more valid conclusion. By using the likelihood metric, we are able to estimate the best data-generating model, which is not possible using only adherence rates.

Of course, readers who firmly believe that choice outcomes are all deterministic and predictable are certainly free to use adherence rates as their favored metric for modeling homogeneous data. But if there is any possibility that this is not the case, then likelihood-based methods should be favored.

References

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: cognitive abilities and information processing. *Journal of Experimental Psychol*- *ogy: General*, *117*(3), 288. doi: 10.1037/0096-3445.117 .3.288

- Alvarez, G. A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, 15(2), 106–111. doi: 10.1111/j.0963-7214.2004 .01502006.x
- Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, 89(4), 369–406. doi: 10.1037/ 0033-295X.89.4.369
- Anderson, J. R., Bothell, D., Lebiere, C., & Matessa, M. (1998a). An integrated theory of list memory. *Journal* of Memory and Language, 38(4), 341–380.
- Anderson, J. R., Bothell, D., Lebiere, C., & Matessa, M. (1998b). An integrated theory of list memory. *Journal* of Memory and Language, 38(4), 341–380. doi: 10.1006/ jmla.1997.2553
- Anderson, J. R., Farrell, R., & Sauers, R. (1984). Learning to program in lisp. *Cognitive Science*, 8(2), 87–129. doi: 10.1016/S0364-0213(84)80013-0
- Ascough, J., Maier, H., Ravalico, J., & Strudley, M. (2008). Future research challenges for incorporation of uncertainty in environmental and ecological decision-making. *Ecological Modelling*, 219(3), 383–399. doi: 10.1016/ j.ecolmodel.2008.07.015
- Atmanspacher, H., & Römer, H. (2012). Order effects in sequential measurements of non-commuting psychological observables. *Journal of Mathematical Psychology*, 56(4), 274–280. doi: 10.1016/j.jmp.2012.06.003
- Audley, R. (1960). A stochastic model for individual choice behavior. *Psychological Review*, 67(1), 1–15. doi: 10.1037/h0046438
- Awh, E., Barton, B., & Vogel, E. K. (2007). Visual working memory represents a fixed number of items regardless of complexity. *Psychological Science*, 18(7), 622–628. doi: 10.1111/j.1467-9280.2007.01949.x
- Axelrod, R. (1973). Schema theory: An information processing model of perception and cognition. *American Political Science Review*, 67(04), 1248–1266. doi: 10.2307/ 1956546
- Baddeley, A. D., & Hitch, G. J. (1974). Working memory. *The psychology of learning and motivation*, *8*, 47–89. doi: 10.1016/S0079-7421(08)60452-1

- Balakrishnan, J., & Ratcliff, R. (1996). Testing models of decision making using confidence ratings in classification. *Journal of Experimental Psychology: Human Perception and Performance*, 22(3), 615–633. doi: 10.1037/ 0096-1523.22.3.615
- Baldo, M. V. C. (2013). Signal detection theory in Hilbert space. *Behavioral and Brain Sciences*, 36(03), 277–278. doi: 10.1017/S0140525X12002865
- Barenco, A., Deutsch, D., Ekert, A., & Jozsa, R. (1995). Conditional quantum dynamics and logic gates. *Physical Review Letters*, 74(20), 4083–4086. doi: 10.1103/ PhysRevLett.74.4083
- Beebe-Center, J. G., Rogers, M., & O'connell, D. (1955). Transmission of information about sucrose and saline solutions through the sense of taste. *The Journal of Psychology*, 39(1), 157–160. doi: 10.1080/00223980.1955 .9916166
- Beilock, S. L., & Carr, T. H. (2001). On the fragility of skilled performance: What governs choking under pressure? *Journal of Experimental Psychology: General*, 130(4), 701–725. doi: 10.1037/0096-3445.130.4.701
- Bengtsson, I., & Zyczkowski, K. (2006). Geometry of quantum states: an introduction to quantum entanglement. Cambridge, UK: Cambridge University Press.
- Bergert, F. B., & Nosofsky, R. M. (2007). A response-time approach to comparing generalized rational and take-thebest models of decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(1), 107–129. doi: 10.1037/0278-7393.33.1.107
- Bliss, J. P., Gilson, R. D., & Deaton, J. E. (1995). Human probability matching behaviour in response to alarms of varying reliability. *Ergonomics*, 38(11), 2300–2312. doi: 10.1080/00140139508925269
- Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2006). The priority heuristic: Making choices without trade-offs. *Psychological Review*, *113*(2), 409–432. doi: 10.1037/ 0033-295X.113.2.409
- Bröder, A. (2000). Assessing the empirical validity of the "take-the-best" heuristic as a model of human probabilistic inference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(5), 1332–1346. doi: 10.1037/0278-7393.26.5.1332
- Bröder, A., & Schiffer, S. (2003). Bayesian strategy assessment in multi-attribute decision making. *Journal of Behavioral Decision Making*, 16(3), 193–213. doi: 10.1002/ bdm.442

- Bröder, A., & Schütz, J. (2009). Recognition rocs are curvilinear—or are they? on premature arguments against the two-high-threshold model of recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(3), 587–606. doi: 10.1037/a0015279
- Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. University of California Press.
- Büning, H. K., & Lettmann, T. (1999). Propositional logic: Deduction and algorithms (Vol. 48). Cambridge University Press.
- Busemeyer, J. R., & Bruza, P. D. (2012). *Quantum models of cognition and decision*. New York, NY: Cambridge University Press.
- Busemeyer, J. R., Matthews, M., & Wang, Z. (2006). A quantum information processing explanation of disjunction effects. In *The 29th annual conference of the cognitive science society and the 5th international conference of cognitive science* (pp. 131–135). Erlbaum.
- Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. *Psychological review*, 100(3), 432–459. doi: 10.1037/0033-295X.100.3.432
- Busemeyer, J. R., Wang, Z., & Lambert-Mogiliansky, A. (2009). Empirical comparison of markov and quantum models of decision making. *Journal of Mathematical Psychology*, 53(5), 423–433. doi: 10.1016/j.jmp.2009.03 .002
- Cave, K. R., & Wolfe, J. M. (1990). Modeling the role of parallel processing in visual search. *Cognitive Psychology*, 22(2), 225–271. doi: 10.1016/0010-0285(90)90017-X
- Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*(1), 55–81. doi: 10.1016/0010-0285(73)90004-2
- Colman, A. M. (2003). Cooperation, psychological game theory, and limitations of rationality in social interaction. *Behavioral and Brain Sciences*, 26(02), 139–153. doi: 10.1017/S0140525X03000050
- Davidson, D., & Marschak, J. (1959). Experimental tests of a stochastic decision theory. *Measurement: Definitions and theories*, 233–69.
- Davis-Stober, C. P., & Brown, N. (2011). A shift in strategy or "error"? strategy classification over multiple stochastic specifications. *Judgment and Decision Making*, *6*(8), 800– 813.

- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological bulletin*, *129*(1), 74–118. doi: 10.1037/0033-2909.129.1.74
- Deutsch, D. (1989). Quantum computational networks. In Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences (Vol. 425, pp. 73– 90). doi: 10.1098/rspa.1989.0099
- Dougherty, M. R., Franco-Watkins, A. M., & Thomas, R. (2008). Psychological plausibility of the theory of probabilistic mental models and the fast and frugal heuristics. *Psychological Review*, 115(1), 199–211. doi: 10.1037/ 0033-295X.115.1.199
- Dougherty, M. R., Gettys, C. F., & Ogden, E. E. (1999). Minerva-dm: A memory processes model for judgments of likelihood. *Psychological Review*, 106(1), 180. doi: 10.1037/0033-295X.106.1.180
- Dreyfus, S. E. (2004). The five-stage model of adult skill acquisition. *Bulletin of Science, Technology & Society,* 24(3), 177–181. doi: 10.1177/0270467604264992
- Egeth, H., Jonides, J., & Wall, S. (1972). Parallel processing of multielement displays. *Cognitive Psychology*, *3*(4), 674–698. doi: 10.1016/0010-0285(72)90026-6
- Eisert, J., & Wilkens, M. (2000). Quantum games. *Journal* of *Modern Optics*, 47(14-15), 2543–2556.
- Fiedler, K. (2010). How to study cognitive decision algorithms: The case of the priority heuristic. *Judgment and Decision Making*, 5(1), 21–32.
- Fontana, G., & Gerrard, B. (2004). A post keynesian theory of decision making under uncertainty. *Journal of Economic Psychology*, 25(5), 619–637. doi: 10.1016/ j.joep.2003.11.001
- Garcia-Retamero, R., Hoffrage, U., & Dieckmann, A. (2007). When one cue is not enough: Combining fast and frugal heuristics with compound cue processing. *The Quarterly Journal of Experimental Psychology*, 60(9), 1197–1215. doi: 10.1080/17470210600937528
- Garner, W. (1953). An informational analysis of absolute judgments of loudness. *Journal of Experimental Psychology*, *46*(5), 373–380. doi: 10.1037/h0063212
- Gigerenzer, G. (2004). Fast and frugal heuristics: The tools of bounded rationality [Book Section]. In D. J. Koehler & N. Harvey (Eds.), *Blackwell handbook of judgment and decision making* (p. 62-88). Oxford, UK: Blackwell.
- Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. Annual Review of Psychology, 62, 451–482. doi: 10.1146/annurev-psych-120709-145346

- Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychological review*, 103(4), 650–669. doi: 10.1037/ 0033-295X.103.4.650
- Gigerenzer, G., Hertwig, R., & Pachur, T. (2011). *Heuristics: The foundations of adaptive behavior*. New York, NY: Oxford University Press, Inc.
- Gigerenzer, G., Hoffrage, U., & Goldstein, D. G. (2008). Fast and frugal heuristics are plausible models of cognition: Reply to Dougherty, Franco-Watkins, and Thomas (2008). *Psychological Review*(1), 230–239. doi: 10.1037/ 0033-295X.115.1.230
- Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, 98(4), 506–528. doi: 10.1037/0033-295X.98.4.506
- Gigerenzer, G., & Todd, P. M. (1999). *Fast and frugal heuristics: The adaptive toolbox.* Oxford University Press.
- Gigerenzer, G., Todd, P. M., & ABC Research Group. (1999). *Simple heuristics that make us smart*. New York, NY: Oxford University Press.
- Glöckner, A. (2007). Does intuition beat fast and frugal heuristics? a systematic empirical analysis. In B. C. Plessner H. & T. Betsch (Eds.), *Intuition in judgment and decision making* (pp. 309–325). New York, NY: Psychology Press.
- Glöckner, A., & Betsch, T. (2008). Do people make decisions under risk based on ignorance? An empirical test of the priority heuristic against cumulative prospect theory. *Organizational Behavior and Human Decision Processes*, 107(1), 75–95. doi: 10.1016/j.obhdp.2008.02.003
- Glöckner, A., & Bröder, A. (2011). Processing of recognition information and additional cues: A model-based analysis of choice, confidence, and response time. *Judgment and Decision Making*, 6(1), 23–42.
- Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, 109(1), 75–90. doi: 10.1037/0033-295X.109 .1.75
- Green, D., & Swets, J. (1966). Signal detection theory and psychophysics. *Psychological Bulletin*, 75, 424–429. doi: 10.1901/jeab.1969.12-475
- Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in*

Cognitive Sciences, *14*, 357–364. doi: 10.1016/j.tics.2010 .05.004

- Hawkins, S. A., & Hastie, R. (1990). Hindsight: Biased judgments of past events after the outcomes are known. *Psychological Bulletin*, *107*(3), 311. doi: 10.1037/0033 -2909.107.3.311
- Hertwig, R., Gigerenzer, G., & Hoffrage, U. (1997). The reiteration effect in hindsight bias. *Psychological Review*, *104*(1), 194–202. doi: 10.1037/0033-295X.104.1.194
- Hertwig, R., Hoffrage, U., & ABC Research Group. (2013). Simple heuristics in a social world. New York, NY: Oxford University Press.
- Hilbig, B. E. (2010). Reconsidering "evidence" for fast-andfrugal heuristics. *Psychonomic Bulletin & Review*, 17(6), 923–930. doi: 10.3758/PBR.17.6.923
- Hoffrage, U., Hertwig, R., & Gigerenzer, G. (2000). Hindsight bias: A by-product of knowledge updating? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(3), 566–581. doi: 10.1037/0278-7393.26.3 .566
- Hogarth, R. M. (1987). Judgement and choice: The psychology of decision. John Wiley & Sons.
- Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology*, 24(1), 1–55. doi: 10.1016/0010-0285(92) 90002-J
- Hogarth, R. M., & Karelaia, N. (2007). Heuristic and linear models of judgment: Matching rules and environments. *Psychological Review*, 114(3), 733–758. doi: 10.1037/0033-295X.114.3.733
- Holyoak, K. J., & Simon, D. (1999). Bidirectional reasoning in decision making by constraint satisfaction. *Journal of Experimental Psychology: General*, 128(1), 3–31. doi: 10.1037/0096-3445.128.1.3
- Hutchinson, J. M., & Gigerenzer, G. (2005). Simple heuristics and rules of thumb: Where psychologists and behavioural biologists might meet. *Behavioural Processes*, 69(2), 97–124. doi: 10.1016/j.beproc.2005.02.019
- Iqbal, A., & Toor, A. (2001). Evolutionarily stable strategies in quantum games. *Physics Letters A*, 280(5), 249–256.
- Jacobs, R. A. (1999). Optimal integration of texture and motion cues to depth. *Vision Research*, *39*(21), 3621–3629. doi: 10.1016/S0042-6989(99)00088-7

- Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics* and biases: The psychology of intuitive judgment (pp. 49– 81). New York, NY: Cambridge University Press.
- Kennedy, M., & Gray, R. D. (1993). Can ecological theory predict the distribution of foraging animals - a critical analysis of experiments on the ideal free distribution. *Oikos*, 68(1), 158–166.
- Kerstholt, J. H., & Jackson, J. L. (1998). Judicial decision making: Order of evidence presentation and availability of background information. *Applied Cognitive Psychology*, *12*(5), 445–454. doi: 10.1002/(SICI)1099-0720(199810) 12:5<445::AID-ACP518>3.0.CO;2-8
- Khrennikov, A. Y., & Haven, E. (2009). Quantum mechanics and violations of the sure-thing principle: The use of probability interference and other concepts. *Journal of Mathematical Psychology*, 53(5), 378–388. doi: 10.1016/j.jmp.2009.01.007
- Klir, G., & Yuan, B. (1995). *Fuzzy sets and fuzzy logic* (Vol. 4). Prentice hall New Jersey.
- Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*(4), 349-370. doi: 10.1037/0096-3445.126.4.349
- Krebs, J. R. (1978). Optimal foraging: Decision rules for predators [Book Section]. In J. R. Krebs & N. B. Davies (Eds.), *Behavioral ecology an evolutionalry approach* (pp. 23–63).
- Kvam, P. D., Busemeyer, J. R., & Lambert-Mogiliansky, A. (2013). An empirical test of type-indeterminacy in the prisoner's dilemma. In *Quantum interaction* (pp. 213– 224). Springer. doi: 10.1007/978-3-642-54943-4-19
- Kvam, P. D., Pleskac, T. J., Yu, S., & Busemeyer, J. R. (2015). Interference effects of choice on confidence: Quantum characteristics of evidence accumulation. *Proceedings of the National Academy of Sciences*, *112*(34), 10645-10650. doi: 10.1073/pnas.1500688112
- Lee, M. D., & Cummins, T. D. (2004). Evidence accumulation in decision making: Unifying the "take the best" and the "rational" models. *Psychonomic Bulletin & Review*, 11(2), 343–352. doi: 10.3758/BF03196581
- Link, S., & Heath, R. (1975). A sequential theory of psychological discrimination. *Psychometrika*, 40(1), 77–105.
- Luan, S., Schooler, L. J., & Gigerenzer, G. (2011). A signaldetection analysis of fast-and-frugal trees. *Psychological Review*, 118(2), 316–338. doi: 10.1037/a0022684

- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281. doi: 10.1038/36846
- Malmberg, K. J. (2002). On the form of ROCs constructed from confidence ratings. *Journal of Experimental Psychol*ogy: *Learning, Memory, and Cognition*, 28(2), 380–387. doi: 10.1037/0278-7393.28.2.380
- Manes, E. G. (1976). *Algebraic theories*. Springer-Verlag New York Inc. doi: 10.1007/978-1-4612-9860-1
- Marewski, J. N., Pohl, R. F., & Vitouch, O. (2011). Recognition-based judgments and decisions: Introduction to the special issue (ii). *Judgment and Decision Making*, 6(1-6), 1.
- Marewski, J. N., Pohl, R. F., Vitouch, O., et al. (2010). Recognition-based judgments and decisions: Introduction to the special issue (vol. 1). *Judgment and Decision Making*, 5(4), 207–215.
- Marewski, J. N., & Schooler, L. J. (2011). Cognitive niches: An ecological model of strategy selection. *Psychological Review*, 118(3), 393–437. doi: 10.1037/a0024143
- Martignon, L., & Hoffrage, U. (2002). Fast, frugal, and fit: Simple heuristics for paired comparison. *Theory and Decision*, 52(1), 29–71. doi: 10.1023/A:1015516217425
- Martignon, L., Katsikopoulos, K. V., & Woike, J. K. (2008). Categorization with limited resources: A family of simple heuristics. *Journal of Mathematical Psychology*, 52(6), 352–361.
- Martignon, L., Vitouch, O., Takezawa, M., & Forster, M. R. (2003). Naive and yet enlightened: From natural frequencies to fast and frugal decision trees [Book Section]. In D. Hardman & L. Macchi (Eds.), *Thinking: Psychological perspective on reasoning, judgment, and decision making* (pp. 189–211). New York, NY: John Wiley & Sons Ltd.
- McLeod, P., Plunkett, K., & Rolls, E. T. (1998). *Introduction to connectionist modelling of cognitive processes*. Oxford University Press.
- Meyer, D. A. (1999). Quantum strategies. *Physical Review Letters*, 82(5), 1052. doi: http://dx.doi.org/10.1103/ PhysRevLett.82.1052
- Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, 63(2), 81–97. doi: 10.1037/h0043158
- Moore, D. W. (2002). Measuring new types of questionorder effects: Additive and subtractive. *The Public Opinion Quarterly*, 66(1), 80–91.

- Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of mathematical Psychology*, 47(1), 90–100. doi: 10.1016/S0022-2496(02)00028-7
- Newell, B. R. (2005). Re-visions of rationality? *Trends in Cognitive Sciences*, 9(1), 11–15. doi: 10.1016/j.tics.2004 .11.005
- Newell, B. R., Rakow, T., Weston, N. J., & Shanks, D. R. (2004). Search strategies in decision making: The success of "success". *Journal of Behavioral Decision Making*, *17*(2), 117–137. doi: 10.1002/bdm.465
- Newell, B. R., & Shanks, D. R. (2003). Take the best or look at the rest? Factors influencing "one-reason" decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29(1), 53–65. doi: 10.1037/ 0278-7393.29.1.53
- Newell, B. R., Weston, N. J., & Shanks, D. R. (2003). Empirical tests of a fast-and-frugal heuristic: Not everyone "takes-the-best". Organizational Behavior and Human Decision Processes, 91(1), 82–96. doi: 10.1016/ S0749-5978(02)00525-3
- Nielsen, M. A., & Chuang, I. L. (2010). *Quantum computation and quantum information*. Cambridge University Press.
- Oppenheimer, D. M. (2003). Not so fast!(and not so frugal!): Rethinking the recognition heuristic. *Cognition*, 90(1), B1–B9. doi: 10.1016/S0010-0277(03)00141-0
- Pachur, T., Bröder, A., & Marewski, J. N. (2008). The recognition heuristic in memory-based inference: Is recognition a non-compensatory cue? *Journal of Behavioral Decision Making*, 21(2), 183–210. doi: 10.1002/bdm.581
- Pachur, T., & Hertwig, R. (2006). On the psychology of the recognition heuristic: retrieval primacy as a key determinant of its use. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(5), 983–1002. doi: 10.1037/0278-7393.32.5.983
- Pachur, T., Todd, P. M., Gigerenzer, G., Schooler, L. J., & Goldstein, D. G. (2011). The recognition heuristic: A review of theory and tests. *Frontiers in Psychology*, 2. doi: 10.3389/fpsyg.2011.00147
- Peirce, C. S. (1880). A boolean algebra with one constant. Collected papers of Charles Sanders Peirce, 4, 1931–35.
- Pleskac, T. J. (2007). A signal detection analysis of the recognition heuristic. *Psychonomic Bulletin & Review*, 14(3), 379–391. doi: 10.3758/BF03194081

- Pleskac, T. J., & Busemeyer, J. R. (2010). Two-stage dynamic signal detection: a theory of choice, decision time, and confidence. *Psychological Review*, 117(3), 864. doi: 10.1037/A0019737
- Pohl, R. F. (2006). Empirical tests of the recognition heuristic. *Journal of Behavioral Decision Making*, 19(3), 251– 271. doi: 10.1002/bdm.522
- Pollack, I. (1952). The information of elementary auditory displays. *The Journal of the Acoustical Society of America*, 24(6), 745–749. doi: 10.1121/1.1906969
- Pothos, E. M., & Busemeyer, J. R. (2009). A quantum probability explanation for violations of 'rational'decision theory. *Proceedings of the Royal Society B: Biological Sciences*, 276(1665), 2171–2178. doi: 10.1098/rspb.2009 .0121
- Qin, J., Raye, C. L., Johnson, M. K., & Mitchell, K. J. (2001). Source ROCs are (typically) curvilinear: Comment on Yonelinas (1999). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27(4), 1110– 1115. doi: 10.1037/0278-7393.27.4.1110
- Reed, M. A. (1989). Speech perception and the discrimination of brief auditory cues in reading disabled children. *Journal of Experimental Child Psychology*, 48(2), 270– 292. doi: 10.1016/0022-0965(89)90006-4
- Rieskamp, J. (2008). The probabilistic nature of preferential choice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(6), 1446–1465. doi: 10.1037/a0013646
- Rieskamp, J., & Otto, P. E. (2006). SSL: A theory of how people learn to select strategies. *Journal of Experimental Psychology: General*, 135(2), 207–236. doi: 10.1037/0096-3445.135.2.207
- Savage, L. J. (1954). *The foundations of statistics*. New York, NY: John Wiley & Sons, Inc.
- Schooler, L. J., & Hertwig, R. (2005). How forgetting aids heuristic inference. *Psychological Review*, 112(3), 610– 628. doi: 10.1037/0033-295X.112.3.610
- Shah, A. K., & Oppenheimer, D. M. (2008). Heuristics made easy: an effort-reduction framework. *Psychological bulletin*, 134(2), 207–222. doi: 10.1037/0033-2909.134.2 .207
- Shanks, D. R., Tunney, R. J., & McCarthy, J. D. (2002). A reexamination of probability matching and rational choice. *Journal of Behavioral Decision Making*, 15(3), 233–250. doi: 10.1002/bdm.413

- Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. University of Illinois Press.
- Simon, H. A. (1959). Theories of decision-making in economics and behavioral science. *The American Economic Review*, 49(3), 253–283.
- Simon, H. A. (1990). Invariants of human behavior. Annual Review of Psychology, 41, 1-19. doi: 10.1146/ annurev.ps.41.020190.000245
- Simsek, O., & Buckmann, M. (2015). Learning-curve analysis of simple decision heuristics. In *The 29th annual conference on neural information processing systems*.
- Smith, J. M. (1974). The theory of games and the evolution of animal conflicts. *Journal of Theoretical Biology*, 47(1), 209–221. doi: 10.1016/0022-5193(74)90110-6
- Smith, J. M. (1982). *Evolution and the theory of games*. Cambridge university press.
- Sowa, J. (1983). Conceptual structures: Information processing in mind and machine. Reading, MA: Addison-Wesley Publishers.
- Taatgen, N. (2005). Modeling parallelization and flexibility improvements in skill acquisition: From dual tasks to complex dynamic skills. *Cognitive Science*, 29(3), 421– 455.
- Tarrataca, L., & Wichert, A. (2012). A quantum production model. *Quantum Information Processing*, 11(1), 189–209. doi: 10.1007/s11128-011-0241-2
- Tervo, D. G., Proskurin, M., Manakov, M., Kabra, M., Vollmer, A., Branson, K., & Karpova, A. Y. (2014). Behavioral variability through stochastic choice and its gating by anterior cingulate cortex. *Cell*, 159(1), 21–32. doi: 10.1016/j.cell.2014.08.037
- Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 285–294.
- Todd, P. M., Gigerenzer, G., & ABC Research Group. (2012). *Ecological rationality: Intelligence in the world*. New York, NY: Oxford University Press.
- Townsend, J. T. (1971). A note on the identifiability of parallel and serial processes. *Perception & Psychophysics*, 10(3), 161–163. doi: 10.3758/BF03205778
- Townsend, J. T. (1990). Serial vs. parallel processing: Sometimes they look like tweedledum and tweedledee but they can (and should) be distinguished. *Psychological Science*, *1*(1), 46–54. doi: 10.1111/j.1467-9280.1990.tb00067.x

- Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling* of elementary psychological processes. CUP Archive.
- Townsend, J. T., Silva, K. M., Spencer-Smith, J., & Wenger, M. J. (2000). Exploring the relations between categorization and decision making with regard to realistic face stimuli. *Pragmatics & Cognition*, 8(1), 83–105. doi: 10.1075/pc.8.1.05tow
- Townsend, J. T., & Wenger, M. J. (2004). The serial-parallel dilemma: A case study in a linkage of theory and method. *Psychonomic Bulletin & Review*, *11*(3), 391–418. doi: 10.3758/BF03196588
- Trueblood, J. S., & Busemeyer, J. R. (2011). A quantum probability account of order effects in inference. *Cognitive Science*, *35*(8), 1518–1552. doi: 10.1111/j.1551-6709 .2011.01197.x
- Tversky, A., & Shafir, E. (1992a). Choice under conflict: The dynamics of deferred decision. *Psychological science*, 3(6), 358–361.
- Tversky, A., & Shafir, E. (1992b). The disjunction effect in choice under uncertainty. *Psychological science*, 3(5), 305–309. doi: 10.1111/j.1467-9280.1992.tb00678.x
- Volz, K. G., Schooler, L. J., Schubotz, R. I., Raab, M., Gigerenzer, G., & von Cramon, D. Y. (2006). Why you think milan is larger than modena: Neural correlates of the recognition heuristic. *Journal of Cognitive Neuroscience*, 18(11), 1924–1936. doi: 10.1162/jocn.2006.18.11.1924
- von Neumann, J. (1955). *Mathematical foundations of quantum mechanics*. Princeton University Press.
- Wang, Z., & Busemeyer, J. (2016a). Comparing quantum versus markov random walk models of judgements measured by rating scales. *Phil. Trans. R. Soc. A*, 374(2058), 20150098. doi: 10.1098/rsta.2015.0098
- Wang, Z., & Busemeyer, J. R. (2013). A quantum question order model supported by empirical tests of an a priori and precise prediction. *Topics in Cognitive Science*, 5(4), 689– 710. doi: 10.1111/tops.12040
- Wang, Z., & Busemeyer, J. R. (2016b). Interference effects of categorization on decision making. *Cognition*, 150, 133–149. doi: 10.1016/j.cognition.2016.01.019
- Wang, Z., Solloway, T., Shiffrin, R. M., & Busemeyer, J. R. (2014). Context effects produced by question orders reveal quantum nature of human judgments. *Proceedings of the National Academy of Sciences*, 111(26), 9431–9436. doi: 10.1073/pnas.1407756111

- Wickelgren, W. A. (1979). Chunking and consolidation: A theoretical synthesis of semantic networks, configuring in conditioning, sr versus cognitive learning, normal forgetting, the amnesic syndrome, and the hippocampal arousal system. *Psychological Review*, 86(1), 44–60. doi: 10.1037/0033-295X.86.1.44
- Yanofsky, N. S., Mannucci, M. A., & Mannucci, M. A. (2008). *Quantum computing for computer scientists* (Vol. 20). New York, NY: Cambridge University Press.
- Yearsley, J. M., & Busemeyer, J. R. (2015). Quantum cognition and decision theories: A tutorial. *Journal of Mathematical Psychology*. doi: doi:10.1016/j.jmp.2015.11.005
- Yonelinas, A. P., & Parks, C. M. (2007). Receiver operating characteristics (ROCs) in recognition memory: A review. *Psychological Bulletin*, 133(5), 800–832. doi: 10.1037/0033-2909.133.5.800