Information and value during multi-attribute learning and decision making

by Cristian G Giron

This thesis was submitted to
The Hong Kong Polytechnic University
in partial fulfillment of the requirements of the
degree of Master of Philosophy in the
Department of Rehabilitation Sciences

August 2020

Supervision by Dr. Bolton KH Chau and Dr. Andy SK Cheng

Table of Contents

Acknowledgements

I would like to express a great deal of appreciation to my chief supervisor, Dr Bolton Chau, for his patient instruction and motivator style of supervision that reinforced my interest in science. Further thanks to Ms Ka Ki Lau and the Lau family for showing me Hong Kong and keeping me from getting lost in the city. And thanks to my colleagues, Mr Michael Woo and Mr Kelvin Law, who were good for discussions about neuroscience, philosophy, and food.

Abstract

Dissecting the computational components of the explore-exploit dilemma is critical to our understanding of how the mind works. A core component of the dilemma is understanding the contexts where option informativeness is either appetitive or irrelevant. In the present thesis, this computational problem was investigated using a novel multi-attribute bandit task and Bayesian model analyses, observing two critical results. First, a behavioral task was used to probe whether informativeness can defined as a quantifiable variable, as opposed to paradigms in the literature that use a categorical operational definition. Indeed, subjects considered this quantifiable definition of informativeness alongside value. Specifically, analyzing the behavioral experiment with traditional statistics demonstrated signature patterns of exploratory behavior that was consistent with the literature. Second, Bayesian modeling allowed further investigation of potential hypotheses underlying these patterns of exploration – namely, the modulatory role of uncertainty in the deliberation of value and informativeness. There are further questions about informativeness to explore, but this thesis presents a means of investigating and exploring this critical construct on more mathematical grounds.

Keywords: value-based decision making, informativeness, learning, computational models

Chapter 1 – Literature Review

Decision neuroscientists endeavor to understand how brains prefer. One contemporary method to accomplish this is by testing algorithmic theories of cognition and behavior to delineate the variable underpinnings of preferential choice (Chau et al., 2018; Farrell & Lewandowski, 2019; Glimcher, 2014; Kriegeskorte and Douglas, 2018; Wilson & Collins, 2019). Some relevant examples include sequential sampling during option valuation (Hunt et al., 2018; Maier et al., 2020; Shenhav et al., 2018), mutual inhibition during option comparison (Chau et al., 2014; Hunt et al., 2014; Wang, 2002), and precision weighting to depend more on reliable information to support our beliefs (Cao et al., 2019; Echeveste et al., 2020; Ernst & Banks, 2002; Meyniel & Dehaene, 2017). Research that furthers efforts to explain and predict value-based choice examine the extent that higher-order cognitive processes control beliefs and behavior (Collins & Frank, 2013; Cools & D'Esposito, 2011; Frank & Fossella, 2011; Kriegeskorte and Douglas, 2018; Lee et al., 2012; Lieder et al., 2018; Ott & Nieder, 2019; Radulescu et al., 2019). The higher-order process investigated here is the weight that perceived informativeness has on preferential choice. To elaborate the research question: do humans deliberate between informativeness and value in multi-attribute environments? If so, can this behavior be expressed algorithmically? The present thesis investigates information-seeking and value-based behavior by subjecting humans to an online, multi-attribute value-based decision-making and learning task. Further, it tests computational models based on Bayes' rule to approximate the algorithms underlying decision making in the task. Bayes' theorem is commonly applied in the decision neuroscience literature to compare human behavior with optimal inference-making under uncertainty (Bach & Dolan, 2012; Farrell & Lewandowski, 2018; Griffiths et al., 2012; Kriegeskorte & Douglas, 2018; O'Reilly, 2013; O'Reilly & Mars, 2015), but the extent that

information-based decision making occurs in multi-attribute environments is sparsely explored. The following sub-chapter briefly describes the theoretical and methodological principles underlying this present thesis. This is followed by sub-chapters on value-based decision making, learning, multi-attribute environments, and informativeness before concluding with a summary on the research gap and questions.

1.1. The thesis and the approach

Three principles from cognitive neuroscience are assumed in this thesis and guide its approach towards the research question about the extent that option informativeness influences preferential choice. The first is the computational theory of the human mind. The theory assumes the brain is an information-processing device wherein neurons transform sensory information into adaptive cognitive and motor behavior (Marr, 1982; Pinker, 2009). The second principle is a consequence of the first. Information-processing is algorithmic in nature, wherein processing events occur systematically. As such, these processes can be represented mathematically. Often termed computational models, these formularizations can be tested with, for examples, psychophysics or with measures of neuronal state and activity (Forstmann et al., 2011; Kriegeskorte and Douglas, 2018; Marr, 1982; O'Reilly and Mars, 2011; Passingham & Rowe, 2014). The third principle follows the first and second, wherein a top-down approach, using computational models to describe higher-order cognitive processes, contributes and advances to our knowledge of mental functions (Chau et al., 2018; Kriegeskorte and Douglas, 2018; Marr, 1982; Passingham & Rowe, 2014). These principles underlie the approach of this thesis. As such, these principles are elaborated on next.

The first is a classic theoretical framework explicitly or implicitly assumed in the cognitive neuroscience literature (Edelman, 2008; Kriegeskorte and Douglas, 2018; Marr, 1982;

O'Reilly and Mars, 2011; Parr et al., 2018; Pinker, 2009). Our ability to perceive the world, identifying its colors and depths, requires astonishing engineering and computational feats that feel effortless but are mathematical ill-posed problems (Edelman, 2008; Marr, 1982; Pinker, 2009). The computational theory of the human mind proposes that brains perform these astonishing feats, solving ill-posed problems, by making computational assumptions and efficiently using information received from the senses to infer the external environment e.g., combining information from multiple senses to infer the state of the environment (Cao et al., 2019; Edelman, 2008; Ernst & Banks, 2002; Meyniel & Dahaene, 2017; O'Reilly & Mars, 2015; Pinker, 2009). An intriguing consequence of this theory is that we do not experience the world but instead experience our brain's inferences or model about our environment, i.e., we are living in the world's shadow (Pinker, 2009). This may explain our surprising susceptibility to optical illusions and behavioral oddities that can be clarified by flaws in algorithmically representation (Cohen et al., 2016; Martens & Wyble, 2010; Pinker, 2009; Watanabe et al., 2018). The theory of an algorithmic brain was made more evident to computer scientists who endeavored to make perceiving machines but found perception a complex problem. Marr (1982) posited that this inspired scientists to treat the brain as an information-processing device in order to understand and engineer perceiving and cognitive machines. Marr proposed three levels for understanding an information-processing device: a computational level describing the purpose of turning informative-inputs to outputs; a representation and algorithmic level describing how inputs are stored and how algorithmic processes transform inputs to outputs; and a hardware level describing the physical implementation of the second level. These laws have made the investigation on *how the brain works* tractable, evidenced by the profundity of the cognitive neuroscience literature (Cohen et al., 2016; D'Esposito & Postle, 2015; Edelman, 2008;

Forstmann et al., 2011; Kriegeskorte and Douglas, 2018; Ma et al., 2014; Marr, 1982; O'Reilly et al., 2012; Passingham & Wise, 2012; Pinker, 2009; Watabe-Uchida et al., 2017).

The second principle posits that computational models, akin to algorithmic hypotheses, identify the variables underlying human behavior and cognition (Farrell & Lewandowski, 2018; Forstmann et al., 2011; Kriegeskorte & Douglas, 2018; Lee, 2013; O'Reilly et al., 2011; Marr, 1982; Radulescu et al., 2019; Rangel & Clithero, 2014; Rescorla & Wagner, 1972; Wilson & Collins, 2019). In their recent overview on this topic, Wilson and Collins (2019) discussed four common applications of computational modeling in psychology and neuroscience: simulation, parameter fitting, model comparison, and latent variable inference. Together, these applications offer a way to examine latent cognitive and behavioral processes and patterns. To elaborate a bit further, analyses of complex systems like the brain benefit from sophisticated methods that can capture subtle but dynamic components of an information processing device beyond what descriptive statistics and linear models can discover. The present thesis approximates information-processing components of decision making by subjecting human subjects to a controlled, multi-attribute learning and choice task. Then, human choice patterns are compared with variants of a computerized chooser, i.e., Bayesian models, whose algorithm and statistical inferences about the environment is known. These variants are then compared, with the winning algorithm determined to be the best algorithmic hypothesis of human behavior. Ultimately, these algorithms can be used to search for brain regions or neuromodulator systems that correspond with the parameters of the winning models. This is a typical workflow in the cognitive neuroscience literature and exemplifies the practicality of the computational theory of mind and computational modeling (Busemeyer et al., 2019; Chau et al., 2018; Edelman, 2008; Farrell & Lewandowski, 2018; Forstmann et al., 2011; Griffiths et al., 2012; Marr, 1982; O'Reilly & Mars,

2015; Parr et al., 2018; Passingham & Rowe, 2014; Passingham & Wise, 2012; Pinker, 2009; Wilson & Collins, 2019). This way of resolving how the brain works is powerful and has been argued to be superior for research compared to invasive neuroscience methods thus far (Niv, Forthcoming).

Kriegeskorte and Douglas (2018) reviewed disparate literatures investigating how the brain works, findings complementary methods and endorsed collaboration between cognitive science, computational neuroscience, and artificial intelligence. To briefly summarize, cognitive science investigates the complex, higher-order cognitive processed with robust and replicable findings (Zwaan et al., 2018). Higher-order cognitive processes include cognitive control, which be examined using computational models, as described above, but suffer from too large a scope, i.e., even the best fitting model is still the product of a myriad of simpler algorithms implemented by neurons that are outside the scope of behavioral and neuroimaging techniques. To ascertain these simpler algorithms, computational neuroscience investigates neuronal implementations, but suffer from too small a scope, i.e., these models describe simple algorithms but do not put these together to produce the sophisticated higher-order cognitive processes. To do so, artificial intelligence applies these simple algorithms to simulate neuronally feasible models to develop computerized intelligence (Kriegeskorte & Douglas, 2018). The present thesis has a place in this collaboration as the hand of cognitive science, because this thesis poses human subjects with a constrained computerized environment so that their cognitive processes can be examined with computational models (Kriegeskorte & Douglas, 2018; Marr, 1982; Niv, Forthcoming; Pinker, 2009). Submitting humans or animals to such environments allow cognitive neuroscientists to investigate a specified mental construct, applied in a broad range of fields including perception, neuroeconomics, and the modern study of human phenomenology

(Cohen et al., 2016; Fiedler et al., 2019; Lau & Rosenthal, 2011). More specifically, these tightly controlled and complex experiments allow cognitive scientists to identify the algorithmic problems resolved by behavior. To paraphrase Kriegeskorte and Douglas (2018) and Marr (1982), we cannot understand what neurons are doing without understanding the problems brains must solve.

In summary, the present work falls under the scope of cognitive neuroscience. It utilizes its theories about the brain to develop behavioral experiment and model analysis to investigate human decision making in multi-attribute environments. In addition to value, the algorithmic nature of information-seeking and its influence on preferential choice are examined. To belabor the research question: does human preferential choice deliberate between information and value – specifically, is informativeness a quantitative component of the valuation process? The following sub-chapters discuss the literature on value-based decision making and learning about simple and multi-attribute options.

1.2. Value-based decision making

Decision neuroscience literature is divided into two psychological constructs: preferential and perceptual decision making, with the former also referred to as *value-based* or *subjective* decision making (Glimcher, 2014). This sub-chapter highlights the differences between these constructs to highlight the core properties of value-based decision making. Value-based decision making is the process of selecting options that are *preferred* because they satisfy the chooser's goals. In real life, it involves selecting between options with various idiosyncratic values. Given this, it is common to define optimal or rational value-base decision-making as behavior that always or more likely selects the option with the highest idiosyncratic values (Busemeyer et al., 2019; Chau et al., 2018; Glimcher, 2014; Lee, 2013; Lee et al., 2012; O'Doherty et al. 2017;

O'Reilly, 2013; Ott & Nieder, 2019). To improve choice predictions, decision neuroscientists investigate the algorithms and neural source of subjective valuations and comparison. This is critical as healthy human decision making cannot be perfectly predicted even when options are objectively and parametrically defined (Chau et al., 2014; Gluth et al., 2018; Hunt et al., 2014; Jamali et al., 2019) or subjective values are obtained before the choice task (Fujiwara et al., 2018; Gluth et al., 2020; Polanía et al., 2019; Voigt et al., 2019). Subjectivity distinguishes preferential from perceptual decision making, wherein perceptual decisions instead depend on the quality of sensory evidence that the brain uses to categorize items (Chalk et al., 2010; Cohen et al., 2016; Kohl et al., 2018). Put another way, preferential decisions depend more on an internal valuation mechanism of evidence whereas perceptual decision depend more on external sensory evidence (Glimcher, 2014; Passingham & Wise, 2012). Clinical cases further demonstrate the qualitative difference between preference and perception. For instance, drug addiction is thought to be the result of a suboptimal or irrational valuation mechanism (Haber & Behrens, 2014; Lee, 2013; Stahl, 2013). Patients may be aware of the aversiveness of their abused substance, but unlike their healthy counterparts, these beliefs do not outweigh the anticipated pleasure from consuming the substance (Calabresi et al., 2007; Cools & Robbins, 2004; Crocket & Fehr, 2014; Dolan & Dayan, 2013; Haber & Behrens, 2014; Lee, 2013; Stahl, 2013). Conversely, clinical examples of perceptual decision deficits abound in psychology. A classic case involves patient DF who suffered lesions to her lateral occipital cortex, leaving her with an inability to identify objects in her view. Interestingly, she is able to make appropriate motor decisions in response to those objects (Goodale et al., 1991; Whitwell et al., 2014). These clinical cases of preferential and perceptual decision-making highlight a qualitative difference between these constructs: the former is dependent on goals and beliefs to guide decisions; the

latter is dependent on sensory evidence to categorize (Dutilh & Rieskamp, 2016; Passingham & Wise, 2012). Further evidence distinguishing preferential and perceptual decision making is found in neuronal activity studies (Dutilh & Rieskamp, 2016; Jamali et al., 2019; Passingham & Wise, 2012). For instance, Jamali and colleagues (2019) performed a rare single-unit recording experiment on human dorsolateral prefrontal cortex (dlPFC), a brain region associated with cognitive control during decision making, working memory, and attention (Chau et al., 2018; Collins & Frank, 2013; Cools & D'Esposito, 2011; D'Esposito & Postle, 2015; Frank et al., 2009; Meyer-Lindenberg et al., 2005; Ott & Nieder, 2019; Radulescu et al., 2019; Seamans & Yang, 2004; Williams et al., 1995). Jamali and colleagues observed that a subset of dlPFC neurons displayed activity patterns that correlated with subjective-decision patterns of patient. Critically, this relationship between dlPFC neurons and decisions was related to subjective but not perceptual decision making (Jamali et al., 2019).

To summarize, decision neuroscience distinguishes preferential and perceptual decision making, implicating separate neural processes and behavioral outcomes. This thesis examines value-based decision making, with this sub-chapter focused on the literature of how values are learned and used to make rational decisions (Busemeyer et al., 2019; Chau et al., 2018; Daw & Tobler, 2014; Dolan & Dayan, 2013; Fiedler et al., 2019; Glimcher, 2014; Lee, 2013; Lee et al., 2012; Mackintosh, 1975; O'Doherty et al. 2017; O'Reilly, 2013; Padoa-Schioppa, 2011; Pearce & Hall, 1980; Rangel & Clithero, 2014; Rescorla & Wagner, 1972; Watabe-Uchida et al., 2017). The decision processes seem to occur in parallel, such that option valuation occurs while options are compared with a response possible upon probing (Busemeyer et al., 2019; Rangel & Clithero, 2014). An easy way to study these processes to observe the deliberation between two, clearly defined options. These *simple choices* allow psychologist, neuroscientists, and economist to

investigate the processes and patterns underlying preference (Fiedler et al., 2019; Rangel & Clithero, 2014). Findings by studies implementing simple-choice tasks describe a neural system composed of the following processes: an accumulative process for option valuation; a competitive inhibitory process for comparison; and a probabilistic process for responses. Two critical points are necessary before these processes are described. First, the literature suggests these stages do not occur sequentially or independently but occur in a distributed, parallel, and hierarchical manner (Hare et al., 2010; Hunt et al., 2014; Sarafyazd & Jazayeri, 2019). Second, this sub-chapter on value-based decision making assumes value learning has already occurred but values must be decoded and integrated during valuation. Learning is discussed in the following sub-chapter.

1.2.1. Valuation: an accumulative process

Evidence from the decision neurosciences suggest that values begin developing immediately after options are presented (Busemeyer et al., 2019; Fiedler et al., 2019; Glimcher, 2014; Lee et al., 2012; Padoa-Schioppa, 2011; Rangel & Clithero, 2014; Ratcliff & McKoon, 2008). The principle is critical in many computational models of decision making, such as *sequential sampling* theories (Busemeyer et al., 2019; Ratcliff & McKoon, 2008). These models propose that evidence for an option's value accumulates up-to some threshold, ultimately producing a decision or belief (Busemeyer et al., 2019; Gluth et al., 2020; Hunt et al., 2018; Juechems et al., 2019; Krajbich et al., 2011; Kohl et al., 2018; Shenhav et al., 2018). To take a recent example, evidence that sequential sampling processes occur in the brain were observed by Hunt and colleagues (2018). Using single-unit recording on monkeys, the authors measured neuronal activity in multiple frontal cortical regions, notably the anterior cingulate cortex (ACC), while these monkeys underwent an attention-controlled decision-making task. ACC is implicated

in several functional roles of value-based decision making and learning including, broadly speaking, uncertainty and belief updating about posed options (Behrens et al., 2007; Busemeyer et al., 2019; Chau et al., 2018; Fouragnan et al., 2019; Haber & Behrens, 2014; Hunt et al., 2018; Juechems et al., 2019; Muller et al., 2019; Shenhav et al., 2018). Hunt and colleagues steered monkey attention by controlling the sequence of required eye fixations onto specific attributes of the options displayed. Thereafter, monkeys were free to sample attributes before decision making. Controlling saccades this way had the effect of controlling the sequence of information sampling because, as studies show (Gluth et al., 2020; Krajbich et al., 2011), the rate that value information for an option accumulates faster with greater attention, with gaze serving as an approximation to attention (Fiedler et al., 2019; Glimcher, 2014). The authors observed that ACC activity was the strongest predictor of valuation beliefs relative to the current best option, i.e., whether the monkey hypothesized that the current best option was supported or rejected by new information through saccades. The finding reaffirms sequential sampling theories because neurons representing beliefs about values were updated with evidence, characteristic of accumulation-to-threshold processes and replicating human brain studies implicating the frontal cortices with valuation processes (Busemeyer et al., 2019; Chau et al., 2018; Haber & Behrens, 2014; Haber & Knutson, 2010; Passingham & Wise, 2012).

In summary, the valuation process begins at choice onset: once options are presented, the brain begins collecting information about value. Further, the process is accumulative, such that value develops as a function of attention time. Evidence for ACC and the effect of new evidence on behavior and neurons was described, but this is not to say the ACC is an independent processing unit for belief updating – the lateral intraparietal region also has an extensive literature implicating it to evidence accumulation (Beck et al., 2008; Dorris & Glimcher, 2004;

Lee et al., 2012; Louie et al., 2011; Passingham & Wise, 2012; Wang, 2002). Indeed, complex cognitive processes like valuation are highly distributed in the brain. To understand this process, it is necessary to identify the algorithms implemented by the brain, as opposed to only localizing neural correlates (Marr, 1982; Passingham & Rowe, 2014). This is true for the comparison process. In the same study, Hunt and colleagues (2018) measured the neuronal activity of the orbitofrontal cortex (OFC), finding that only monkey OFC neurons encoded dynamic value comparison as opposed to valuation. That is, comparison was specific to monkey OFC and was not observed in the other brain regions the authors investigated, i.e., ACC nor dlPFC (Hunt et al., 2018).

1.2.2. Comparison: an inhibitory process

It is worth repeating that although valuation and comparison are described separately, evidence suggests these processes occur in a parallel, hierarchical, and distributed manner. Put succinctly: the evidence suggests that brains prefer while they evaluate. As mentioned above, Hunt and colleagues (2018) observed comparison signals in monkey OFC, consistent with the decision neuroscience literature pointing to the role of human ventromedial prefrontal cortex (vmPFC) in value comparison (Bartra et al., 2013; Chau et al., 2018; Chau et al., 2014; Chau et al., 2020; Glimcher, 2014; Haber & Behrens, 2014; Haber & Knutson, 2010; Jocham et al., 2012; Lee, 2013; Lee et al., 2012; Levy & Glimcher, 2011; Rangel & Clithero, 2014; Rouault et al., 2019; Shiner et al., 2012; Voigt et al., 2019; Walton et al., 2015). Whereas value development was described as a sequential and accumulative process, preference development is described here as a competitive and inhibitory process (Chau et al., 2014; Fouragnan et al., 2019; Hunt et al., 2018; Wang, 2002). Two theories about value comparison are described next: the first presents an empirically supported theory that demonstrates calling value comparison an

inhibitory process as well as being an evidence supporting research into the brain as an information processing device; and the second feature explains how value comparison between different domains of value is possible.

The work by Chau and colleagues (2014) is notable for predicting behavior and neural activity with a modified biophysical model (Passingham & Rowe, 2014). The model originally was designed by Wang (2002) as an inhibitory value comparison process termed mutual inhibition (**Figure 1a**). During simulations, a high valued option and low valued option were represented by a notional pool of neurons (respectively, 'P_{HV}' and 'P_{LV}'). The biophysical model posits that each pool of neurons receives excitatory input in proportion to the value of the option each pool represents. Each pool then delivers excitatory input to a shared, inhibitory pool ('P_i') that simultaneously feeds back inhibitory inputs to each pool. The result is mutual inhibition, wherein each pool, 'P_{HV}' and 'P_{LV}', excites an inhibitory pool that in turn inhibits the original pool and the opposing pool of neurons (i.e., via 'P_{HV}' inhibits 'P_{LV}' and vice versa via 'P_i'; **Figure 1a**). In addition to mediating this competition, this inhibitory component was included to control the excitation of the neural pools in the network (Brunel & Wang, 2001). 'P_{HV}' receives greater excitatory inputs because it represents a higher valued option. It is therefore better suited to withstand inhibitory inputs from 'P_i' compared to 'P_{LV}', which receives less excitatory input. Consequently, 'P_{HV}' has a higher signal-to-noise ratio than 'P_{LV}', granting these neurons a greater chance be selected for the decision process (Chau et al., 2014). Note the result of this competition only makes a decision *more likely*, implying the lack of determinism in decision making. There is stochasticity in decision making even when decisions are easy, e.g., when HV – LV is large. This is discussed in the next sub-chapter. Here, it suffices to point-out how value

comparison is modeled in the brain as an inhibitory process. The result is a biophysical model of the competitive, comparison process termed *mutual inhibition* (Chau et al., 2014; Wang, 2002).

Chau and colleagues (2014) extended this model of value comparison to observe the consequences of a third, distracting option (**Figure 1b**). This model predicted a nonintuitive effect of the value of distracting option D, with representative neuronal pool 'P_D', on the population activities of 'P_{HV}' and 'P_{LV}'. Specifically, because of the influence of 'D' on the inhibitory interneuron pool, the model predicted diminished value comparison accuracy with lower values of 'D'. First, the authors confirmed that it was easier for subjects to choose 'HV' when its value was greater than 'LV'. Next, the authors observed that it was harder to choose 'HV' when the distracting option 'D' was smaller (Chau et al., 2014). Their biophysical model predicted these results (**Figure 1b**): a smaller distractor value resulted in decreased interneuron inhibition in the model, followed by decreased control of the signal-to-noise ratio when the network continued to run the competition between 'P_{HV}' and 'P_{LV}' when the distractor option was revealed as unavailable (see Brunel & Wang, 2001, for a discussion on the inclusion of an inhibitory interneuron component in the recurrent network model). Chau and colleagues also observed neural correlates of the value comparison in the vmPFC and medial intraparietal sulcus using functional magnetic resonance imaging (fMRI). Notably, vmPFC signals were also weaker when 'D' was lower, and the signal strength scaled with subjective decision accuracy of the human subjects (Chau et al., 2014). The finding that the vmPFC is implicated in value comparison is consistent with the literature (Bartra et al., 2013; Glimcher, 2014; Haber & Behrens, 2014; Haber & Knutson, 2010). The result above suggests that lapses in healthy decision making may be due to algorithmic limitations of a neurobiological implementation of value comparison (Chau et al., 2014).

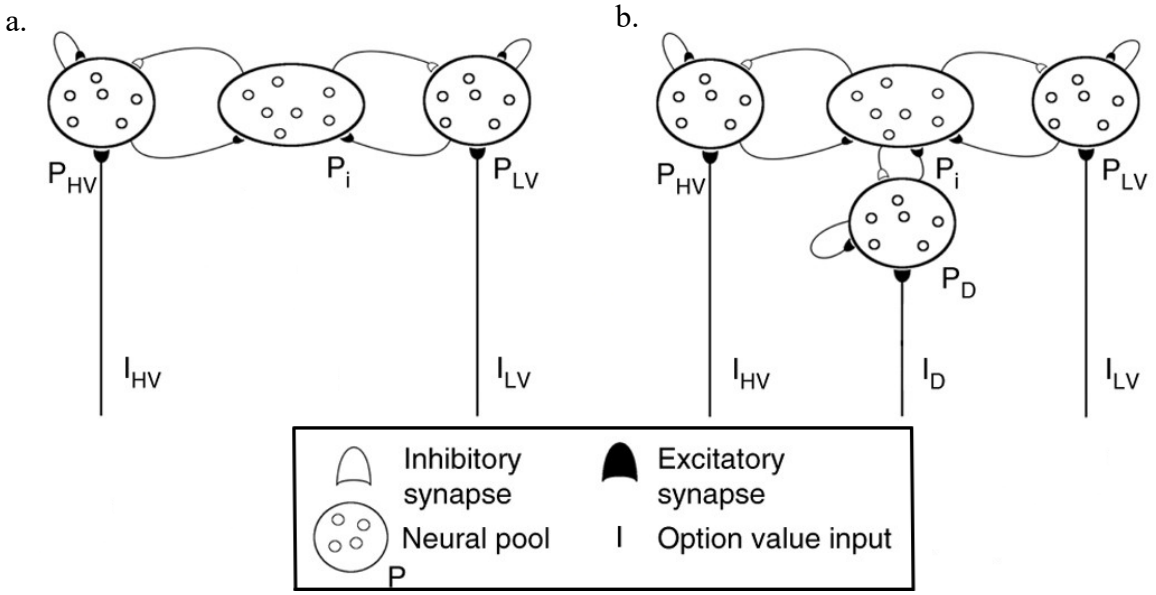


Figure 1. *A biophysical model of value comparison. The value of the high-valued option is referred to as 'HV' and the alternative, low-valued option is referred to as 'LV'.* ***(a)*** *A feasible model for comparison of two options with an inhibitory mediator.* ***(b)*** *An extension of Wang's (2002) model to include a third option. The figure is modified and adapted from Chau and colleagues (2014).*

The second but critical feature of value comparison in the brain is the concept of a *common neural currency* (Bartra et al., 2013; Levy & Glimcher, 2011; Rangel & Clithero, 2014). Humans can deliberate between options that are categorically different, such as choosing to keep writing instead of going out for an egg tart. Empirical evidence supporting the concept of a common neural currency was observed by Levy and Glimcher (2011). The authors tested subjects on a binary decision-making task where options were either money or food. The authors found distinct neural mechanisms for the valuation of money versus food: money was processed in the posterior cingulate cortex whereas food preference was processed in the hypothalamus. Levy and Glimcher next tested subjects on a version of the task where money was compared with food on each trial, allowing the authors to estimate how much money a certain amount of food is worth to each subject. The authors observed that the vmPFC represented these scaled valuations, further implicating the region with using a common neural currency that allowed human subjects to make decisions comparing categorically different options (Levy & Glimcher,

2011). Findings such as these and Chau and colleagues (2014) implicate the vmPFC as critical in the algorithms involving value comparison (Bartra et al., 2013; Behrens et al., 2007; Chau et al., 2018; Chau et al., 2014; Chau et al., 2020; Glimcher, 2014; Haber & Behrens, 2014; Haber & Knutson, 2010; Jocham et al., 2012; Lee, 2013; Lee et al., 2012; Levy & Glimcher, 2011; Muller et al., 2019; Padoa-Schioppa, 2011; Rangel & Clithero, 2014; Rouault et al., 2019; Shiner et al., 2012; Voigt et al., 2019; Walton et al., 2015). With options evaluated and compared, preferences are developed. Then, decision makers can respond in accordance with these preferences. Though the quality of this accordance is not straightforward.

1.2.3. Response: a probabilistic process

If one were to form an opinion about human decision making by examining computational models, decision making would look like a probabilistically process. Here, valuation functions simply influence decision probabilities, e.g., high value options have a higher probability of being chosen (Farrell & Lewandowski, 2018; Lee et al., 2012; Wilson & Collins, 2019). To predict human choice behavior, studies use statistical tools to estimate the probability of a choice. Notably, these statistical tools artificially introduce noise in order match human behavior (Ballard et al., 2018; Chau et al., 2015; Collins & Frank, 2012; Hunt et al., 2012; Leong et al., 2017; Niv et al., 2015; Shiner et al., 2012). The softmax function one such example, often used as the chooser for many types of computational models:

$$p_x(EV_x) = \frac{\exp\left(\frac{EV_x}{T}\right)}{\sum_{i=1}^{n} \exp\left(\frac{EV_i}{T}\right)} \qquad \text{[Equation 1]}$$

value an option is weighed by a subjective temperature parameter, *T*. This parameter is used to estimate the level of stochasticity of the chooser, such that small values of *T* indicate precise decision and greater values indicate noisy decisions (Wilson & Collins, 2019), e.g., when *T* is

small, the expected value of an option's is more consequential, such that a greater expected value produces a greater probability of being chosen. Behaviorally, if a chooser can successfully discriminate between the choice values and accurately chooses high value options, then the chooser would be described by a small temperature, *T*. Conversely, temperature is large when decisions are random and indiscriminate of value. In this manner, the softmax function is an algorithmic expression for determining a chooser's sensitive to value and the results probability they would choose according to values. A critical caveat is that the cause of this noise is typically not discussed. A recent paper from Polanía and colleagues (2019) discussed this issue and described the potential source of decision noise observed in value-based decision making. Their behavioral evidence suggested that noise, introduced in the conversion of comparison to response processes, is a by-product of Bayesian encoding and decoding of value representations by the brain. This would not be surprising as many features of the aforementioned decision processes are approximated by Bayesian models (Griffiths et al., 2012), such as the integration of evidence in accordance to precision during perception (Echeveste et al., 2020; Meyniel & Dehaene, 2017), findings from studies that manipulate information in the environment in way that targets the prior or likelihood components of a Bayesian observer and accordingly biasing human decisions (Chalk et al., 2010; Ting et al., 2015), and studies finding neural and neuromodulator correlates of uncertainty estimated by Bayesian inference (Muller et al., 2019). What is critical from these studies is that probabilistic algorithms are used to described the response process during decision making: responses appear to be a probabilistic process.

1.3. Learning

Studies on value learning investigate how options and outcomes become associated. A well-accepted and intuitive theory is that value associations are the consequence of experience

and transfer (Daw & Tobler, 2014; Lee et al., 2012; Radulescu et al., 2019; Rangel & Clithero, 2014;). In this view, values are learned through trial-and error, wherein a chooser observes the consequences of choosing some option and then assigns value to the relevant attributes of that option. In the present thesis, the algorithmic hypothesis that estimates a developing value in accordance with trial-to-trial observations is called a *value functions* (Lee, 2013; Lee et al., 2012). Two types of value functions discussed in this sub-chapter are Rescorla-Wagner learning and Bayesian inference.

The Rescorla-Wagner model captures iterative learning of animals and humans with successful simulations of behavior (Le Pelley et al., 2016; Rescorla & Wagner, 1972) neurobiological evidence of its implementation in the brain (Chau et al., 2018; Frank & Fossella, 2011; Lee et al., 2012; Radulescu et al., 2019; Schultz et al., 1997; Watabe-Uchida et al., 2017). Even so, the learning model has critical limitations: it lacks a clear measure of certainty and the model struggles to learn if options are not specifically defined (Radulescu et al., 2019; Sutton & Barto, 2018). This problem has led new work studying how the brain assesses task structures to associate with outcomes (Collins & Frank, 2013; Radulescu et al., 2019). It is discussed here to present iterative learning. Value functions can also be Bayesian, capable of learning by trial-and-error while further estimating the level of uncertainty in its estimates (Bach & Dolan, 2012; Courville et al., 2006; Griffiths et al., 2010; Kriegeskorte & Douglas, 2018; O'Reilly, 2013; Muller et al., 2019). Though Bayesian models have neurobiological support (Ballard et al., 2018; Beck et al., 2008; Behrens et al., 2007; Chalk et al., 2010; Meyniel & Dahaene, 2017; Muller et al., 2019; O'Reilly et al., 2012; Ting et al., 2015) it is controversial to claim that brain deploys Bayesian inference or something approximating its model components computations (Bowers &

Davis, 2012a/2012b). Learning is discussed in this thesis by describing the literature on the aforementioned learning algorithms.

1.3.1. By trial-and-error

A simple example of iterative learning, or learning by trial-and-error, is classical conditioning, wherein learning occurs when an initially neutral stimulus, e.g., a light or sound, is followed by a reward that elicits an unconditioned response, e.g., food can be a reward that elicits salivation. If the pairing between the neutral stimulus occurs enough times, the stimulus and food become associated, invoking(Mackintosh, 1975; Rescorla & Wagner, 1972; Pearce & Hall, 1980). This stimulus-reward trial-to-trial association is thought to occur arithmetically in the brain, with the amount of reward expected, the *expected value*, updated iteratively by a reward prediction error (RPE) by the dopaminergic system (Watabe-Uchida et al., 2017).

The RPE is simple formulated as (Rescorla & Wagner, 1975):

$$RPE = \alpha * (received - expected) \quad \text{[Equation 2]}$$

In words, the RPE is the contrast between the outcome *received* after selecting an option with the outcome that was *expected*. This difference is subsequently weighted by a learning rate, α, which determines the extent that the resultant RPE updates our expectations in the future. It can take values between 0 and 1. A value closer to 1 describes a fast and adaptive learner whereas a value closer to 0 describes an inflexible or slow learner (Behrens et al., 2007; Courville et al., 2006). This simple arithmetic can capture many phenomena of learning (Le Pelley et al., 2016; Pearce & Mackintosh, 2010; Rescorla & Wagner, 1975; Watabe-Uchida et al., 2017). Learning itself occurs when the expected value of options in the world, or the option-outcome association, is updated by the RPE:

$$V_{t+1} = V_t + RPE \quad \text{[Equation 3]}$$

The expected value of an option at time *t*, V_t, is updated by RPE (Equation 2), producing the expected value for the next time, $t + 1$. The simplest example of this type of learning is Pavlov's dog; when the dog was initially neutral to the bell, it believed the bell was neither rewarding nor aversive: the expected value, V_t, for the bell equaled zero. Once food was delivered at the same time that the bell was rung, the value of the received reward was greater than what was expected:

$$\mathrm{R}PE = \alpha * (1 - 0) > 0 \qquad \text{[Equation 4]}$$

Assuming the dog's learning rate, α,was greater than 0, and the received reward can be quantified (for simplicity, it is assumed here the reward was a positive magnitude of 1), then the RPE should be a positive value with a magnitude indicating the amount of learning. In the literature, what is being learned can be due to the level of surprise or the extent that the bell predicted the food (Mackintosh, 1975; Pearce & Hall, 1980). In any case, the weight of RPE magnitude is determined by the learning rate and the value difference between the reward and what was expected:

$$V_{t+1} = (0 + (RPE > 0)) > 0 \qquad \text{[Equation 5]}$$

With several trials of bell-food pairings, the expected value will gradually update towards the actual food value and the dog will have completely associated the bell with the food (Rescorla & Wagner, 1975). Conversely, when the value of the received reward is smaller than expected, such as when no food is delivered but was expected, then the RPE is negative with a magnitude proportional to the extent of disappointment (Daw & Tobler, 2013; Lee et al., 2012; Mackintosh, 1975; Pearce & Hall, 1980; Pearce & Mackintosh, 2010; Rescorla & Wagner, 1972; Watabe-Uchida et al., 2017). Another feature of this learning algorithm is flexibility of the learning rate, which is influenced by the volatility of the environment (Behrens et al., 2007; Courville et al., 2006; O'Reilly, 2013). For example, in a volatile environment where option-reward associations

change without warning, the learning rate will be high (near to 1) so that RPEs are large and the learner can adapt to its volatile environment. Conversely, in stable environments where option-reward associations do not change or change predictable, the learning rate will be low (near to 0) so that learning does not dramatically change after one unexpected outcome. An adaptive learner will possess a flexible learning rate so that RPEs effect decision making in accordance with the learning environment.

In the context of modeling of human behavior during experimentation, as in psychophysics task, the Rescorla-Wagner model computes the RPE after feedback is presented. The properties of the chosen option and feedback is then used to approximate the subject's learning experience:

$$V_{t+1} = V_t + \alpha * \ (feedback\ value - V_t) \qquad \text{[Equation 6]}$$

As in **Equation 3**, notation V_t on the right-hand side denotes the present belief on trial *t* with an initial value on trial one depending on the task, i.e., a value during a state of ignorance. For learning models, this value is updated after each trial by the RPE to iteratively compute V_{t+1}, which is then equal to V_t value for the next trial. This learning mechanism is simple and does well to explain behavior, but there are learning phenomenon that it cannot explain. For one, the model lacks an internal estimate of uncertainty. For instance, volatile environments induce uncertainty, which has been observed to affect the learning rate; but the algorithmic process of this flexible behavior is not included in the Rescorla-Wagner model (Radulescu et al., 2019). Additionally, the model does not clearly capture certain learning phenomenon, such as blocking effects, wherein a stimulus with a strong association disrupts learning about a second stimulus (Daw & Tobler, 2013; Rescorla & Wagner, 1972). Rescorla & Wagner (1972) proposed the effect occurred because the second, new stimulus does not provide new information about the

environment that the first, initially informative stimulus already provides. This implies that all the objects and attributes in the environment are assigned an RPE for each option outcome. Algorithmically, Rescorla & Wagner posited an RPE-operation for each attribute in the value function of V_{t+1}. A criticism of this proposed learning strategy is that it is computationally expensive (Cohen et al., 2016; Leong et al., 2017; Niv et al., 2015; Radulescu et al., 2019; Sutton & Barto, 2018), so-called the 'curse of dimensionality.' In the next sub-chapter, it is presented as a reason for the need to study learning and decision strategies in multi-attribute environments.

The neural systems of learning and decision making are highly distributed and hierarchical (Busemeyer et al., 2019; Chau et al., 2018; Passingham & Wise, 2012; Radulescu et al., 2019), further involving neuromodulator systems controlling neuronal signaling and connectivity dynamics in the prefrontal cortex, with dopamine being particularly relevant in the study of RPE-type of learning (Cools & D'Esposito, 2011; Crockett & Fehr, 2014; Frank & Fossella, 2011; Lee, 2013; Ott & Nieder, 2019; Watabe-Uchida et al., 2017). But due to methodological limitations to study prefrontal dopamine, the role of this neuromodulator is typically inferred from non-invasive techniques (Cools & D'Esposito, 2011; Chau et al., 2018; Ott & Nieder, 2019). An essential process for learning as described above is neuroplasticity (Calabresi et al., 2007; Cools & D'Esposito, 2011; Frank & Fossella, 2011; Ott & Nieder, 2019; Radulescu et al., 2019; Shen et al., 2008; van Schouwenburg et al., 2010). Notably, plastic frontostriatal connections are thought to underlie learning mechanisms described by RPE-based algorithms (Ballard et al., 2018; Reynolds et al., 2001; Shen et al., 2008; van Schouwenburg et al., 2010). RPE computations are specifically attributed to dopamine neurons, which project from the midbrain to the prefrontal and striatal regions of the brain that utilize the RPE signal for learning, decision making, attention, representation learning, and motor control (Chau et al.,

2018; Collins & Frank, 2013; Cools & D'Esposito, 2011; Cools & Robbins, 2004; Crocket & Fehr, 2014; Dolan & Dayan, 2013; Eshel et al., 2015; Eshel et al., 2016; Frank & Fossella, 2011; Haber & Behrens, 2014; Haber & Knutson, 2011; Ott & Nieder, 2019; Radulescu et al., 2019; Shiner et al., 2012; Stahl, 2013; Watabe-Uchida et al., 2017).

Interestingly, arithmetic RPE patterns have been observed in the firing patterns of midbrain dopamine neurons (Eshel et al., 2015; Schultz et al., 1997). For example, the rate of dopamine neuron activity in the nucleus accumbens increases with positive model estimates of RPEs, i.e., when the actual outcome is greater than expected in the task. Conversely, dopamine neuron activity drops below baseline after a negative RPE, when an outcome is lower than expected (Roitman et al., 2008; Shen et al., 2008; Schultz et al., 1997; Wickens et al., 2007). Further, this dopaminergic signal corresponds to the magnitude of the RPE: a highly unexpected reward produces a relatively high dopaminergic response (Fiorillo et al., 2003; Matsumoto & Hikosaka, 2009). Anatomically, greater release of dopamine results in faster learning corresponding with increased potentiation of frontostriatal connections (Lee et al., 2012; Watabe-Uchida et al., 2017). In a pharmacological study on mice, Reynolds and colleagues (2001) observed frontostriatal potentiation when the dopamine producing neurons in the substantia nigra pars compacta were stimulated. Reynolds and colleagues then applied a dopamine D1-receptor antagonist and saw potentiation decrease after the same stimulation procedure, demonstrating that midbrain dopamine has a role in frontostriatal potentiation (Reynolds et al., 2001). This RPE-based plasticity in mice is also observed in humans and monkeys, suggesting a generalized learning mechanism across species (Matsumoto & Hikosaka, 2009; Passingham & Wise, 2012; Radulescu et al., 2019; Redgrave et al., 1999; van Schouwenburg et al., 2010). For instance, van Schouwenburg and colleagues (2010) conducted an fMRI study with human subjects to test

whether the basal ganglia had a role in delivering top-down signals that emphasize goal-relevant stimuli in the posterior visual processing regions. The authors observed striatal activity modulated PFC functional connectivity with posterior sensory regions, suggesting that the striatum tells the PFC when to switch top-down inhibitory signals to emphasize goal-relevant stimuli (van Schouwenburg et al., 2010). Further studies suggest that biasing attention in this way enhances frontostriatal potentiation or depression that is in turn mediated by the dopamine system (Calabresi et al., 2007; Cools & D'Esposito, 2011; Crocket et al., 2013; Ott & Nieder, 2019; Redgrave et al., 1999).

Modern psychopharmacological methods also support dopamine's role in trial-and-error learning, as the Rescorla-Wagner seems to capture (Burke et al., 2018; Cools et al., 2009; Cools & D'Esposito, 2011; Chau et al., 2018; Crockett & Fehr, 2013). For instance, positron emission transmission scanning can track injected dopamine transmission in the nervous system during behavioral experiments (Cools et al., 2008; Liu et al., 2017;). In other studies, asking patients who regularly take dopaminergic medication to withhold their treatment allows researchers to use neuroimaging while these patients perform experiments, then later be compared with healthier brains, to observe manifestations of dopaminergic deficits in humans (Pine et al., 2010; Shiner et al., 2012; Tost et al., 2009). Less invasive methods involve the use of dopaminergic genes whose dispositions influence the transmission of prefrontal or striatal dopamine critical for trial-and-error learning (Chau et al., 2018; Cools & D'Esposito, 2011; Doll et al., 2011; Doll et al., 2016; Elton et al., 2017; Filla et al., 2018; Frank et al., 2007; Frank et al., 2009; Frank & Fossella, 2011; Gao et al., 2016; Gershman & Tzovaras, 2018; Meyer-Lindenberg et al., 2005; Meyer-Lindenberg et al., 2007; Persson & Stenfors, 2018; Slifstein et al., 2008; Tost et al.,

2009). Though this can be described by Rescorla-Wagner, it is not claimed here that this algorithm is exactly performed by the dopaminergic system and the network it mediates.

Trial-and-error learning as observed behaviorally has a neural bases: the brain generates RPE signals and delivers these learning signals to the PFC and striatum where option values store and compared (Ballard et al., 2018; Behrens et al., 2007; Chau et al., 2018; Daw & Tobler, 2013; Eshel et al., 2015; Eshel et al., 2016; Glimcher, 2013; Haber & Behrens, 2014; Lee et al., 2012; Leong et al., 2017; Ott & Nieder, 2019; Pearce & Mackintosh, 2010; Schultz et al., 1997; Watabe-Uchida et al., 2017). Modern treatments for post-traumatic stress disorders and phobias involve reversal learning, a method similar to the one above for dissociating the sound of a bell from food. In the case of post-traumatic stress disorder, exposure therapies are used, which teach patients that a conditioned stimulus no longer predicts some traumatic event, an effort to dissociate a learned association (Bryant & Nickerson, 2013). As well, combining these theories with clinical applications, decoded neurofeedback was developed, used dissociate fear responses (Chiba et al., 2019). This all suggests that learning values by trial-and-outcome has a neural basis with RPE-based algorithms. However, Rescorla-Wagner model is just algorithmic hypothesis that approximates this type of learning.

1.3.2. By Bayesian inference

An alternate value function, applied in this thesis (**Chapter 3 – Bayesian Models**), is trial-and-error learning by Bayesian inference. Models based on Bayesian inference provide several advantages over the Rescorla-Wagner model. First, because Bayes' theory is a mathematically optimal way to combine new with old information (Griffiths et al., 2012; Stone, 2013), these models are used in the literature to simulate optimal behavior in tasks requiring inference. These simulations can be compared with human performance to identify where

behavior strays or is comparable with optimized behavior, such as learning and value-based decision making (Bach & Dolan, 2012; Collins & Frank, 2013; Frank et al., 2009; Kriegeskorte & Douglas, 2018; Meyniel & Dahaene, 2017; Muller et al., 2019). A second reason for choosing Bayesian inference over the Rescorla-Wagner model is the former accounts for the imprecision that is involved during behavior in uncertain situations. This can also be simulated and compared with human behavior to gain insight about the computational problems that brains must solve (Marr, 1982). Incidentally, neurons appear to approximate Bayesian computations (Beck et al. 2008; Chalk et al., 2010; Echeveste et al., 2020; Ting et al., 2015), although, this interpretation of is debatable (Bowers & Davis, 2012a/2012b; Griffiths et al., 2012). This sub-chapter describes the application of Bayesian inference in human learning.

Bayesian inference is a statistical method for computing the probability of a hypothesis give new and old evidence (Griffiths et al., 2012; Farrell & Lewandowsky, 2018; O'Reilly et al., 2012; Stone, 2013). This can be expressed as:

$$p(hypothesis \mid data) \propto p(data \mid hypothesis) * \ p(hypothesis)$$

Posterior Likelihood Prior

or symbolically as:

$$p(hyp|\ x) \propto p(x|hyp) * p(hyp) \qquad \text{[Equation 7]}$$

Posterior Likelihood Prior

The above term is Bayes' Factor, wherein hyp symbolizes some hypothesis under investigation and x represents the data observed. It should be noted that Bayes' factor is a simplification of Bayes' theorem. In the above formulation, the left-hand side is proportionate to the right and are not necessarily equivalent. Bayes' theorem is more constraining:

$$p(hyp|x) = \frac{p(x|hyp) * \ p(hyp)}{p(x)} \qquad \text{[Equation 8]}$$

Where hyp and x still represent some hypothesis and some data, respectively, but now, this term is normalized by the probability of the data, termed the marginal likelihood (Farrell & Lewandowski, 2018; Stone, 2013). As this value is the same for the space of hypotheses under study, it is not necessary for analyses that compare posterior distributions, the left-hand side of **Equation 7** and **8**, which is the case in present thesis (Farrell & Lewandowski, 2018; Stone, 2013). In any case, Bayes' theorem is a mathematically optimal method for probabilistically integrating new with past data to infer the most probable hypothesis under study (Farrell & Lewandowski, 2018; Griffiths et al., 2012; O'Reilly & Mars, 2015; Stone, 2013). In order to compute a posterior probability distribution, the prior and likelihood terms must be specified for every hypothesis and possible outcome, sometimes called the statespace or parameter space (Farrell & Lewandowsky, 2018; O'Reilly, 2013; Wilson et al., 2010). This is an important limitation, because it means the Bayesian model cannot make inferences about an undefined state, a point influencing the way Bayesian models were design in this thesis.

Bayes' theorem's capability to compute the probability distribution of many possible outcomes is used hypothesis testing in wide range of scientific disciplines including psychology, neuroscience and engineering (Kriegeskorte & Douglas, 2018; O'Reilly et al., 2012; O'Reilly & Mars, 2015; Stone, 2013). Relevant here, cognitive neuroscientists use Bayesian models to approximate human inference making (Griffiths et al., 2012; O'Reilly et al., 2012; Wilson et al., 2010). Bayesian estimates the probability of a hypothesis, among a statespace of hypotheses, being true. Hinging beliefs on the highest probability in this distribution or the weighted mean of this distribution would be optimal inference and utilized in the literature on perceptual and preferential decision making and learning (Chalk et al., 2010; Griffiths et al., 2012; Meyniel & Dahaene, 2017; Muller et al., 2019; Nassar et al., 2010; O'Reilly et al., 2012; O'Reilly & Mars,

2015; Ting et al., 2015). This procedure is herein called Bayesian learning, and looks like the diagram in **Figure 2** adapted from Chalk and colleagues (2010).

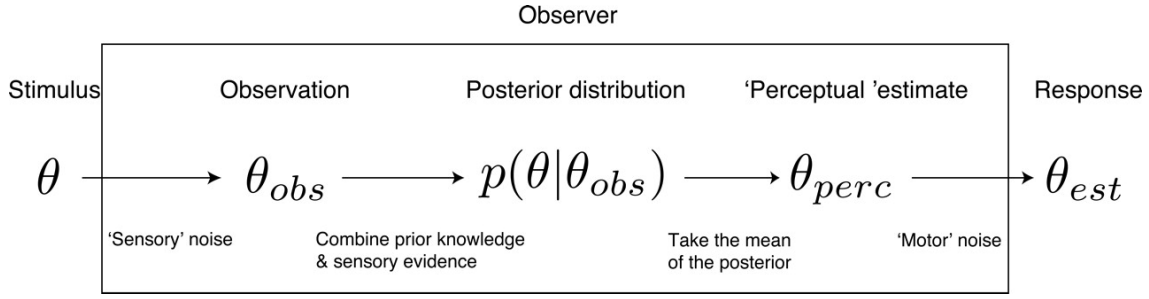


***Figure 2**. Modeling Bayesian learning. Adapted here from Chalk and colleagues (2010) because the diagram summarizes the components of a Bayesian learner in the cognitive neuroscience literature.*

The **Figure 2** diagram by Chalk and colleagues (2010) concisely summarizes Bayesian modeling. It is used here to describe the theory and literature underlying Bayesian modeling. At the start of this process, observers witness evidence or data, notated as the stimulus θ in the diagram. Because of inherent noise in perception, the potential for optimal behavior in response to θ is already diminished (Chalk et al., 2010; Polanía et al., 2019). Consequently, the observation used for Bayesian inference is an approximation of θ, termed θ_{obs}. It is this observation, as opposed to the objective θ, that is used in Bayes' factor (**Equation 7**) – note, as mentioned in the introduction of this chapter, humans do not experience the objective external world, but instead experience its inference (Pinker, 2009). Continuing, the likelihood of seeing θ_{obs} given a hypothesis is weighted by the probability of the hypothesis being true, or the prior in **Equation 7**. This product gives the posterior, which is the probability for hypothesis given θ_{obs}. This can be done for all possible hypotheses in the statespace; obtaining products for each gives a posterior probability distribution. The hypothesis with the highest probability, the maximum a posteriori (MAP), can be used to produce a 'perceptual estimate' (θ_{perc}) of θ_{obs}. With this estimate, the Bayesian observer can make Bayesian optimal decisions. But to do so, θ_{perc} needs to be converted into a motor response. For reasons analogous to noise assumed in the

softmax function for predicting value-based decisions in the preceding chapter, decoding θ_{perc} into a motor response further contributes to the diminished human performance (Chalk et al., 2010; Polanía et al., 2019). Ultimately, this process returns an observable response based on θ_{est}. Such a response would be approximately Bayes-optimal, given environmental, sensory and motor imprecision (Bach & Dolan, 2012; Chalk et al., 2010; Griffiths et al., 2012; Muller et al., 2019; O'Reilly, 2013; Polanía et al., 2019; Ting et al., 2015).

Whether human brains actually implement Bayes' theorem may not be falsifiable with current research methods (Bowers & Davis, 2012ab), but empirical evidence has shown that behavioral and neural patterns are approximately Bayesian (Bach et al., 2012; Beck et al., 2008; Chalk et al., 2010; Collins & Frank, 2013; Courville et al., 2006; Frank et al., 2009; Griffiths et al., 2012; Körding & Wolpert, 2006; Kriegeskorte & Douglas, 2018; O'Reilly et al., 2012; O'Reilly, 2013; O'Reilly & Mars, 2015; Meyniel & Dahaene, 2017; Muller et al., 2019; Nassar et al., 2010; Parr et al., 2018; Polanía et al., 2019; Ting et al., 2015; Wilson et al., 2010). A few studies from perceptual- and preferential- decision making paradigms using Bayesian inference are described. Note, perceptual paradigms manipulate stimulus salience, whereas preferential paradigms typically control stimulus salience and instead manipulate stimuli values (Chau et al., 2018; Fiedler et al., 2019; Glimcher, 2014; Rangel & Clithero, 2014). As such, **Figure 2** applies to value-based decision making as it does in the perceptual literature. Instead of the θ_{obs} in **Figure 2** representing an observed visual stimulus, θ_{obs} can represent an observed value, as opposed to the observed percept, in preferential paradigms (Polanía et al., 2019). Indeed, this idea that perceptual and preferential processing shares a common neural network is has a neural basis in the literature on cognitive maps (Boccara et al., 2019; Chau et al., 2018; Constantinescu et al., 2016).

The first example to discuss evidence on the efficacy of applying Bayesian inference with a neural basis is the work by Chalk and colleagues (2010). The authors tested a set of computational models to find the best algorithmic description of human behavior in a perceptual motion detection task. The authors manipulated the schedule of stimuli motion to induce priors strong enough to cause hallucinated motion in the predicted direction when no stimulus was presented in the task. Additionally, these induced priors improved motion detection when stimulus motion coincided with the prior. This behavior was best described by a Bayesian model that based decisions on an iteratively updated prior. Further, subjects were not aware their expectations were manipulated. This suggested that these Bayesian-like processes were not under conscious control but instead the outcome of internal processing, such as many cognitive biases (Nisbett & Wilson, 1977). Results such as these support the theory that human perception is approximately Bayesian (Cao et al., 2018; Chalk et al., 2010; Ernst et al., 2002; Meyniel et al., 2017).

The following examples describe relevant studies that used Bayesian inference do investigate value-based decision-making . Ting and colleagues (2015) submitted human participants to a lottery computerized task outside and then inside an fMRI. In the task, the weight of evidence, or the likelihood, was manipulated on each decision trial. The authors initially observed that posterior estimates computed by a normative Bayesian model (like the diagram in **Figure 2**) predicted behavior in the task. In other words, using the model estimates of the posterior as the subjects' value functions was better at predicting behavior than only using prior or likelihood estimates. Next, the authors found that the mean of a posterior distribution, a summary-statistic also used in Chalk and colleagues (2010), had representation in the medial prefrontal cortex. Notably, the mean of the prior distribution and likelihood also had distinct

representation in this region. Finally, the optimal integration of these prior and likelihood patterns corresponded with Bayesian model estimates – analogous to precision weighting (Ernst & Banks, 2002; Meyniel & Dehaene, 2017). In addition to using Bayesian inference for hypothesis testing, researchers also investigate the extent that uncertainty plays during decision making. A study by Muller and colleagues (2019) subjected human participants to a different lottery game while undergoing fMRI. Critically, the game's underlying statistics, i.e., the probability that an option is associated with a high reward versus a low reward, were volatile and changed without warning. The authors tested whether uncertainty, as measured by an ideal Bayesian observer model, could describe neuronal activations without obvious changes in the choice preference of subjects. That is, without any overt behavioral effects, the authors investigated whether neuroimaging estimates of uncertainty corresponded with Bayesian model estimates of uncertainty. The authors reported localized brain regions with activation patterns tracked by uncertainty estimates of the Bayesian model. This was further supported by a physiological measure: when an observer was presented surprising information i.e., when θ_{obs} highly strayed from the posterior (**Equation 7**), then pupils dilated. This dilation pattern of surprise is an indirect index of increased uncertainty and noradrenaline release (Preuschoff et al., 2011). Muller and colleagues observed that pupil dilations corresponded to an increase in their Bayesian model's estimated uncertainty, each further corresponded with neuronal activation signatures of uncertainty.

To summarize, the use of Bayesian modeling in cognitive neuroscience has been a fruitful method to investigate cognition and behavior because they help explain and predict choice patterns as well as neuronal representation (Ballard et al., 2018; Behrens et al., 2007; Cao et al., 2019; Courville et al., 2006; Ernst & Banks, 2002; Frank et al., 2009; Meyniel et al., 2017;

Muller et al., 2019; Polanía et al., 2019). Specifically, Bayesian model estimates have been used to estimate the influence of subjective parameters and latent cognitive variables underlying learning and decision making (Bach et al., 2012; Courville et al., 2006; Körding & Wolpert, 2006; Kriegeskorte & Douglas, 2018; O'Reilly, 2013; O'Reilly & Mars, 2015; Parr et al., 2018) such as adaptive learning (Behrens et al., 2007), explorative versus exploitative decision making (Frank et al., 2009; Gershman & Tsovaraz, 2018; Muller et al., 2019), and optimal integration of sensory information (Cao et al., 2019; Ernst & Banks, 2002; Meyniel & Dehaene, 2017). Though the literature and present thesis cannot conclude whether human brains are indeed Bayesian, the evidence suggests, at least, that we cannot reject the Bayesian observer theory of the brain. In any case, the literature demonstrates the utility of Bayesian modeling in cognition and behavior research for hypothesis testing and tracking uncertainty. Therefore, these models are used here to examine information seeking during multi-attribute learning and decision making.

1.4. Multi-attribute learning and decision making

A limitation of the Rescorla-Wagner model described above was that it does well to learn about singularly defined objects but encounters a learning problem in multi-attribute environments. But unlike these models, human learners can identify relevant or predictive attributes for preferred outcomes with limited computational resources (Ballard et al., 2018; Leong et al., 2017; Niv et al., 2015). To illustrate this multi-attribute problem, imagine playing a "Spot the Difference" game wherein you are asked to find differences between two photos of the same messy environment. At first glance, the photos look the same. To find differences, you could try holding the photos away from you so you can conduct a parallel search for obvious discrepancies. This likely will not work, since peripheral vision computes a summary of the things surrounding your focal point and fails to identify subtleties in the environment (Cohen et

al., 2016). Conversely, you could try looking at each a subset of pixels your eye can process and compare each subset to the myriad of other subsets. Obviously, this is not how we choose to solve this game nor is it feasible given limits to our computational resources (Cohen et al., 2016; D'Esposito & Postle, 2015; Edelman, 2008; Ma et al., 2014; Wilhelm et al., 2013). Nevertheless, humans navigate the real world with relative ease and learn about its multi-attribute objects – this sub-chapter discusses the literature on strategy to resolve this computational problem.

Multi-attribute environments are more realistic representations of the world and pose an algorithmic learning challenge that is often neglected by decision neuroscientists (Radulescu et al., 2019). In learning, that problem is the computational complexity of the environment. Another reason to consider attribute-based analyses is empirical evidence that humans can generalize learned attribute associations to novel stimuli (Rangel & Clithero, 2014). So-called the *attribute integration model of subjective value computation*, this theory posits that options are valuated according to their attributes as in the following:

$$\text{Option Value} = \sum (\beta_i * \text{attribute}_i) \qquad \text{[Equation 9]}$$

In the formulation, the value of an option is equal to the sum of all its attributes (i.e., their salience or amount) and their corresponding weights, β, indicating an attribute's relevance to value which depends on the priorities of the chooser. The β may vary across subjects, giving rise to idiosyncratic differences during preferential decision making. The critical implication of this equation is that option values depend on specific attributes (Hare et al., 2009; Hunt et al., 2014; Maier et al., 2020). For example, Hare and colleagues (2009) conducted a decision-making experiment with two accentuated attributes: food taste and healthiness. Based on subject performance in the behavioral task, they grouped human subjects according to their ability to self-control and avoid taste over healthiness: groups included those who had self-control, those

who lacked self-control, and a neutral group who did not outweigh taste over healthiness and vice versa. Using fMRI, the authors found that preference was represented in the vmPFC whether or not subjects exercised self-control. Notably, the nature of this value correlation within vmPFC was group dependent: if subjects had self-control, vmPFC activations correlated with health and taste preferences; but if subjects were not in the self-control group, vmPFC activations only correlated with taste preference (Hare et al., 2009). In a separate fMRI study examining the role of attributes during value-based decision making, Hunt and colleagues (2014) found that intraparietal sulcus (IPS) activity correlated with value comparison between attributes while the dorsal medial prefrontal cortex (dmPFC) correlated with comparison of the whole option. In other words, the IPS activations correlated with the individual products of **Equation 9** before summing, while the dmPFC correlated with its final sum of **Equation 9**. Hunt and colleagues posited that this was evidence of a hierarchical process of valuation, wherein different regions of the brain make value comparisons from varying points of view, e.g., dmPFC compares integrated option values and the IPS compares option attribute values; which comparison influences decision making depends on the chooser's goals (Collins & Frank, 2013; Hunt et al., 2014; Sarafyazd & Jazayeri, 2019). However, as discussed in the trial-and-error sub-chapter on Rescorla-Wagner model, an attribute-based valuation processor would be overwhelmed by the myriad of attributes that compose the real world, so-called the "curse of dimensionality" (Sutton & Barto, 2018).

1.4.1. "Curse of dimensionality"

A simple solution to the "curse of dimensionality" is to make use of strategies that reduce resource demands (Radulescu et al., 2019; Sutton & Barto, 2018). To this end, selective attention can be used to bias learning to a subset of attributes in search of the attributes composing

predictive options in novel and complex environments (Collins & Frank, 2012; Cools & D'Esposito, 2011; Le Pelley et al., 2016; Leong et al., 2017; Niv et al., 2015; Ott & Nieder 2019; Radulescu et al., 2019). This process of learning relevant option structures, also called the *structure representation,* for associative learning is termed *representation learning* and is learned by trial-and-error (Radulescu et al., 2019). This learning depends on selective attention, which attends to different structure representations in the environment in search of most stable and predictive of preferred outcomes. Attention to a structure representation facilitates learning about that structure, so that choice outcomes become associated with it, at the cost of learning little to nothing about other possible unattended structure representations (Leong et al., 2017; Niv et al., 2015; Radulescu et al., 2019). If the structure representation is highly predictive of a rewarding or aversive outcome, then it will be associated with those outcomes. Conversely, if the structure representation is a poor predictor of a rewarding or aversive outcome, then selective attention will test a different structure representation in search of a better combination of attributes in the environment (Ballard et al., 2017; Le Pelley et al., 2016; Leong et al., 2017; Mackintosh, 1975; Niv et al., 2015; Pearce & Mackintosh, 2010; Radulescu et al., 2019). This process of representation learning is thought to take advantage of RPE-like signals described in the preceding section, but learning is about the predictiveness of structure representations facilitated by selective attention (Ballard et al., 2018; van Schouwenburg et al., 2010). For instance, van Schouwenburg and colleagues (2010) used fMRI to monitored the sensory regions of human subjects a task requiring attention-switching. Their stimuli overlapped face and place images, and rules in the task asked subjects to adaptively focus their attention on either stimulus; as expected, focus on one stimulus resulted in the failure to notice when the other updated. Activation enhancements in the sensory regions were goal-dependent, e.g., activations in the

fusiform face gyrus increased when faces needed to be attended and likewise for place stimuli and parahippocampal place area (van Schouwenburg et al., 2010). The critical finding was that the basal ganglia, a region including the striatum which encodes RPE signals and connections to the prefrontal cortex (Ballard et al., 2018; Chau et al., 2018; Haber & Behrens, 2014; Haber & Knutson, 2010), signaled attention-switching. These findings demonstrate that a lack of attention to stimuli results in a failure to track changes of that stimuli in the environment. Likewise, learning requires attention, but its exact role is under debate.

1.4.2. Opposing models of attention

Although the effects of selective attention have been investigated during multiple timepoints of learning and decision making (Gluth et al., 2018; Hunt et al., 2018; Maier et al., 2020) the nature of attentional bias is debatable (Pearce & Mackintosh, 2010; Radulescu et al., 2019). Though attention itself is not investigated in this report, the characteristic behaviors of the type of attention deployed are examined in this thesis, i.e., exploitation and exploration. These two behaviors are characteristic of two highly contested attentional theories concerned with where attention is biased during learning and decision making. The first theory is attributed to Mackintosh (1975), wherein attention is biased towards the most predictive stimuli, as discussed above. This can be observed in exploitative value-based decision making predicted by the Rescorla-Wagner model: the learner seeks and attends the most rewarding stimuli, then the learner will exploit this option by continuously choosing and attending to it at the cost of exploring and learning about other options. The second theory was described by Pearce & Hall (1980), proposing that attention is biased towards stimuli that a learner is the most uncertain about. This can be observed when we choose options that are the most informative, regardless of reward amount (Le Pelley et al., 2016; Pearce & Mackintosh, 2010). Another point of debate

regarding selective attention is whether unconscious or conscious control of attention is more important for learning (Le Pelly et al., 2016). A distinction between unconscious and conscious attentional control is during their competitive deployment underlying exploitative versus exploratory decision making (Dolan & Dayan, 2013; Frank et al., 2009; Gershman & Tzovaras, 2018; Le Pelley et al., 2016; Ott & Nieder, 2019). In their review of attention and learning theories, Le Pelley and colleagues (2016) suggested that studies investigating attentional bias should be designed to encourage both types of attention, unconscious and conscious, so as to assess the dissociable roles of these processes in learning and decision making. The assumption here is that human behavior implements both types of bias, deploying each according to their needs, priorities, and rules posed by the environment (Beesley et al. 2015; Dolan & Dayan, 2013; Wang et al., 2018). Though attention is not measure in the task of this thesis, these two types of attentional bias can be traced by observing response patterns: exploitative decision making is defined by value-based decision making (Mackintosh, 1975) and exploratory decision making yields a bias for informative decisions (Pearce & Hall, 1980). The deployment of these choice strategies varies with task requirements, such as environmental volatility or neuromodulator state (Behrens et al., 2007; Chau et al., 2018; Collins & Frank, 2013; Cools & D'Esposito, 2011; Dolan & Dayan, 2013; Findling et al., 2019; Frank & Fossela, 2011; Frank et al., 2009; Gershman & Tsovaraz, 2018; Le Pelley et al., 2016; O'Reilly, 2013; Ott & Nieder, 2019; Pearce & Mackintosh, 2010; Radulescu et al., 2019; Sarafyazd & Jazayeri, 2019; Trudel et al., 2020; Walton et al., 2015; Warren et al., 2017; Wilson et al., 2014; Zajkowski et al., 2017), but with accurate analyses, such as an appropriate computational model, these choice strategies can be made explicit.

1.5. Informativeness

Exploratory behavior is defined here as a bias towards seeking information in order to reduce uncertainty (Wilson et al., 2020). In other words, option informativeness is specifically the quality that reduces uncertainty, identifiable by the decision making patterns it invokes: when subjects enter a state of high uncertainty, as when the predictiveness of our options are unknown or volatile (Behrens et al., 2007; Muller et al., 2019), then informativeness should contribute an 'information bonus' to the value of an option (Wilson et al., 2014); conversely, when the predictiveness of our options are known or consistent, then the informativeness of an option should not affect the value of an option. Consequently, option informativeness is critical when learning values for decision making. That is, decisions under uncertainty should consider option informativeness to facilitate learning so that value-based decisions are more accurate in the long-run (Walton et al., 2015). This need is explicit in environments or contexts where outcome values must be learned by trial-and-error (Behrens et al., 2007; Frank et al., 2009; Le Pelley et al., 2016; Wilson et al., 2020). To be clear, the appetitive effects of informativeness on decision making is related to associative learning theories of attention described in the previous sub-chapter. According to Pearson and Hall (1980), attention is biased towards the options whose outcomes we are most uncertain about; likewise, during times of high uncertainty, preferential decision making is biased towards option informativeness. In the present thesis, only decision making was analyzed – there is no gaze detection or other attention estimates to report in this present thesis. Therefore, the literature on uncertainty and informativeness on decision making is detailed further in this sub-chapter. The clarification of this definition and pattern are critical to effectively communicate the research gap and questions in the final sub-chapter.

1.5.1. A quantitative variable

As mentioned above, the informativeness of an option reduces uncertainty, and this value is different from the expected reward of an option. An intuitive and technical example are described next (**Figure 3**). The utility of informativeness is explicit in trial-and-error learning. Say you become ill and are in search of a good-enough treatment. You are initially posed two options: Treatment L (the pill in **Figure 3a**) and Treatment R (the syringe in **Figure 3a**). You first try Treatment R but find your symptoms worsen. You fortunately survive for another opportunity to choose between Treatments L and R. The bad experience of Treatment R causes some avoidance, inclining you towards Treatment L in this new opportunity. This treatment makes your symptoms better, thus enabling your future inclination for the good Treatment L. With more experience, you find Treatment L is dependably good. Later, you are able to choose between a new Treatment I and the good Treatment L (**Figure 3b)**: should you explore this unknown treatment or continue exploiting the known treatment? This choice conflict is termed the explore-exploit dilemma. The reasons for choosing the unknown option may be intuitive: perhaps you are curious about its outcome; or perhaps you feel you are testing your luck with the known option and should consider alternatives. In any case, choosing the new option is akin to choosing to explore, this choice reduces your uncertainty about the treatments - a fact that may not be immediately worthwhile given the potential aversive outcome but knowing the qualities of your available treatments is useful in the long term. This dilemma can be represented in more technical terms (**Figure 3b**).

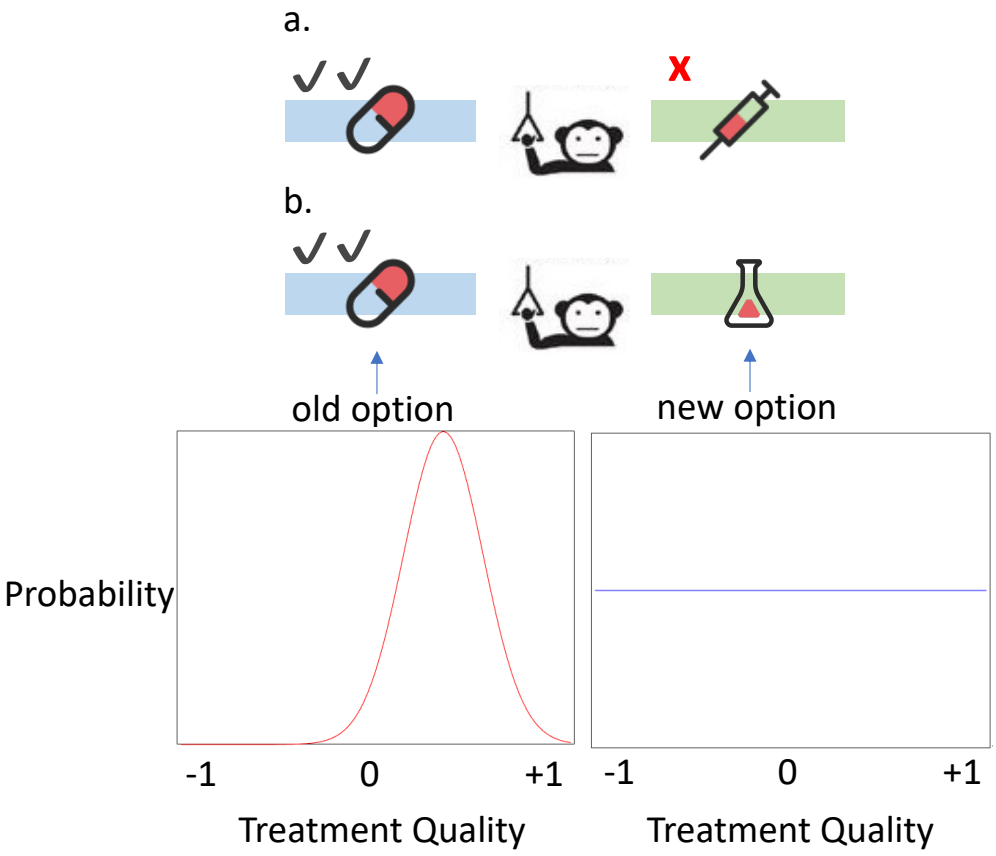


Figure 3. *The role of informativeness in trial-and-error learning.* ***(a)*** *Brief outcome history of two treatments. A checkmark indicates a positive, desired outcome; a red x indicates a negative, aversive outcome.* ***(b)*** *Top: one new treatment, vial, is presented with the old, learned treatment, pill. Below: during decision making, subjects must consider the expected reward (estimated, e.g., using the mean or mode of the probability distribution) and informativeness (the reduction of spread or width of the probability distribution).*

Estimates of value and uncertainty during and trial-and-error learning can be tracked with methods from probability theory (Farrell & Lewandowsky, 2018; O'Reilly & Mars, 2015; Stone, 2013). First, the expected reward from an option and the certainty in that expectation can be represented as a probability distribution: that is, the distribution after plotting the probability for each element in the statespace of possible hypothesis (**Figure 3b**; also discussed in the sub-chapter **1.3.2. Bayesian inference**). To elaborate, options can be described as having two features essential during any kind of decision making: the first feature is whether outcome quality is appetitive or aversive; the second feature is whether this quality is constant or volatile

(Behrens et al., 2007; Muller et al., 2019). The first feature can be represented by the mean or mode of a probability distribution (**Figure 4**). To clarify this idea, imagine seeing a new treatment option as discussed above – its treatment quality and reliability is unknown. This state of ignorance can be represented by a *uniform distribution*: all possible treatment qualities have an equal probability of being the true value (Farrell & Lewandowsky, 2018; Stone, 2013; **Figure 4**, **Trial 1**). That is, the new treatment has an equal probability of being healing as it does of being poisonous. Decisions based on this state of ignorance, a uniform distribution, equate to guesses . Once the new option is chosen, its outcome is observed for the first time; let's say this outcome is healing or appetitive. The update from the **Trial 1** to **Trial 2** plot in **Figure 4** demonstrates a learning effect: the probability distribution of possible treatment values converges under the observed value, with the plot now suggesting that the new treatment is good but with much uncertainty. The latter suggestion is a consequence of the width of distribution. After a second experience, **Trial 3**, the probability distribution converges further, indicating that the learner is more confident that the treatment quality is the observed value and more likely to occur again in future outcomes. **Trial 4** demonstrates the effects that the same treatment would produce if its outcome was suddenly poisonous or aversive. This makes the learner uncertain about the treatments true value, indicated by a diverging distribution, and that the true value of the treatment may be aversive, indicated by a mode shifted towards the left side of zero. This thesis proposes that, while value affects the mode or mean of these probability distributions that reward based decisions are based on, information seeking is based on the width of these distributions. In this latter property, when the width is relatively large, informativeness is very appetitive and manifests as exploratory behavior that appears like expected reward is neglected; conversely, when the width is small, the informativeness of an option is negligible and expected reward is

critical, which manifests as exploitative behavior (Dolan & Dayan, 2013; Walton et al. 2015; Wilson et al., 2020).

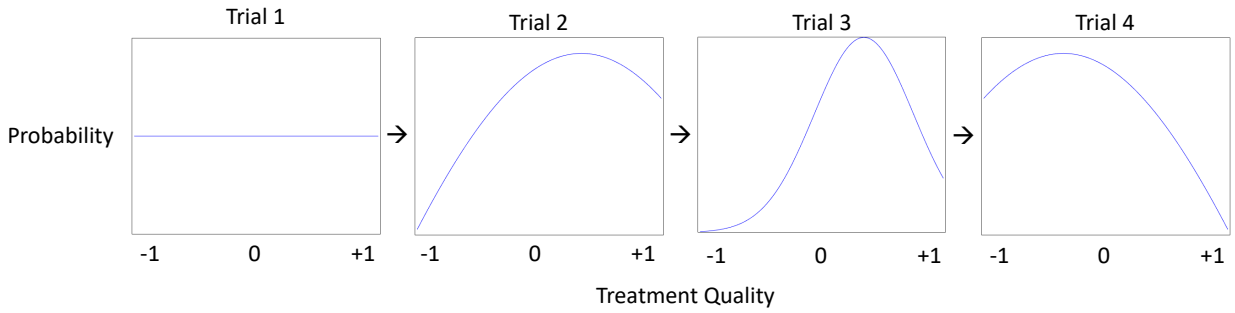


Figure 4. *Trial-and-learning learning represented as updating probability distributions, or priors, of the estimated treatment quality.*

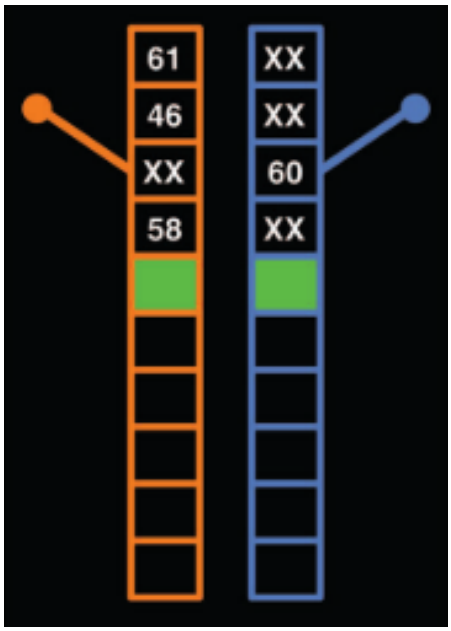
This method of tracking uncertainty, using the distribution widths of learned variable as those in **Figure 4** and their relationship with information seeking, was implemented in Trudel and colleagues (2020). The authors posed subjects with two options on each trial, each defined by their predictiveness of a rewarding outcome. Here, subjects were required to learn option predictiveness by trial-and-error and use this knowledge to make decisions that maximize total reward in the task. Using Bayes' rule to track estimates of predictiveness and uncertainty of the subjects, and fMRI to localize BOLD correlates in the brain, the authors identified three behavioral phases underlying patterns of value-based decision-making during learning: an exploratory phase, an exploitative phase, and a transition phase between these two phases. Critically, the exploratory phase was not random choice behavior, instead appearing strategic. Subjects preferred more uncertain options, i.e., less precise and predictive of rewarding outcomes, during early periods of a block. After a few trials, subjects switched to an exploitative phase where they preferred options that were more predictive of rewards. BOLD imaging patterns shifted with behavioral phases, described by the authors as a polarity change. The vmPFC, previously observed to be correlated with preference (Chau et al., 2014; Voigt et al.,

2019), was positively correlated in Trudel and colleagues (2020) with uncertainty during the exploratory phase - suggesting a preference for uncertainty. Then, vmPFC BOLD activity was negatively correlated with uncertainty in later exploitative phase – suggesting an aversion for uncertainty. Finally, this behavior was affected by the number of opportunities subjects had to make a choice before the predictiveness of the options changed. More known opportunities with the same options invoked longer exploratory periods (Trudel et al., 2020), replicating the conditions wherein overt and strategic exploratory behavior was manipulated (Warren et al., 2017; Wilson et al., 2014). More on the application of Bayesian modeling is discussed in **Chapter 3 – "Bayesian Models"**; here it is sufficient to observe the use of these models in the literature to examine information seeking. The operation definition of this construct in the relevant literature is discussed in the next sub-chapter.

1.5.2. Categorical definitions

While informativeness is understood as a quantitative variable in the literature, it is typically operationally defined as a categorical variable. For instance, Wilson and colleagues (2014) produced a fairly recent study comparing the deployment of information-based versus value-based decision making. Their findings and paradigm inspired investigations using neuroimaging (Trudel et al., 2020), brain stimulation (Zajkowski et al., 2017) and psychopharmacology (Warren et al., 2017); these papers have additionally replicated the information seeking patterns observed by Wilson and colleagues. However, although Wilson and colleagues designed a clever experiment and paradigm, in essence, the authors binarized informativeness (**Figure 5**). For instance, their task controlled the amount of information subjects had about each option, done by manipulating the number of times subjects observed the outcomes of each possible choice. The left diagram in **Figure 5** displays a sample trial in the task

as described by Wilson and colleagues (2014), used here to described their operational definition of informativeness. In the trial, subjects had to choose between a left-red slot machine or a right-blue slot machine. Wilson and colleagues controlled how much information subjects obtained about each slot machine by forcing their first four choices. In the sample trial in **Figure 5**, subjects were forced to choose the left-red three times and then the right-blue once. Then, on the fifth trial, subjects were free to choose either one of the two options: if subjects chose the option with the least amount of information, the right-blue option with only a single observation, then subjects were described as having made an informative choice (Wilson et al., 2014). Else, subjects were said to have made a value-based decision by ignoring the high uncertainty of the right option for the certainty of the left. Several aspects of the task were controlled, including the number of prospective options and the average reward amounts of their options – properties of the choice environment that influence the deployment of exploratory behavior. The critical point here is that a decision was defined as being information-seeking or exploitative. This operational definition of decision making was unable to capture the quantitative nature of informativeness discussed in the previous sub-chapter in the present thesis. This was also the case in the authors computational models (**Figure 5**, inside the box on the right-hand side). The value function estimates the value of an option after integrating all relevant variables; in the controlled task in Wilson and colleagues, this involved the amount of reward and an "information bonus", positive for informative options as operationally defined, negative otherwise. This way of operationally defining informativeness, either a choice is or is not due to information seeking, however clever and inspirational in the literature, is problematic and limiting in the study of human exploratory behavior (Findling et al., 2019; Frank et al., 2009; Trudel et al., 2020; Warren et al., 2017; Wilson et al., 2014; Zajkowski et al., 2017).



option value$_i$ = reward$_i$ + **information$_i$**

if an option is more informative,

information$_i$ = 1

else,

information$_i$ = -1

Figure 5. *The behavioral paradigm (left) and a summary of the computational model (right) developed by Wilson and colleagues (2014).*

1.6. Research gap and questions

Several critical points from the cognitive neuroscience literature on value-based decision making and learning in multi-attribute environments were discussed in the literature review. This included evidence of thee hierarchical, parallel, and distributed processes that underlie value-based decision, trial-and-error theories of learning, computational problems, and solutions for learning and decision making in multiple attribute environments, and the current theories of the importance of informativeness attribute of our options. In the present thesis, the design of the behavioral experiment and Bayesian models were based on this research as discussed above. For instance, when option values are known and the environment is predictable, an exploitative decision strategy to maximize rewards will be optimal. But when attributes must be learned, as when the option-outcome associations are unstable, then a chooser should be more tactful, deploying exploratory learning strategies such as prioritizing option informativeness over value. These exploratory strategies must be tempered by exploitative strategies lest the final reward be disappointing. In an optimal chooser, these strategies are balanced in a way to enable learning while maximizing gains (Wilson et al., 2020). The neural and algorithmic theories underlying

learning and decision making were discussed in the preceding literature review as well, with the purpose to demonstrate and neural evidence for the methods used in the behavioral and modeling studies of this thesis.

The research gap under question lies in the definition of informativeness that is sought during exploratory strategies to resolve a decision problem. Specifically, the literature's operational definition of informativeness is categorical, either a decision is or is not an information seeking decision. This definition is rather constraining, as the effects of informativeness may be quantitative, as suggested by probability theory methods used to describe human learning. The research questions are twofold. First, can informativeness sought during exploration Can "informativeness" be captured by a more quantifiable definition? Second, can computational models of decision making be improved by using this new definition of informativeness? The next two chapters describe two methods to investigate the research question: **Chapter 2 – "Behavioral experiment"** describes the computerized environment developed to tease out information- and value-based decision making, examining the extent of deliberation between information and value in a multi-attribute environment; **Chapter 3 – "Bayesian models"** describes computational models to further examine these strategies by comparing the behavioral results with computerized performance.

Chapter 2 – Behavioral Experiment

2.1. Introduction

A behavioral experiment was conducted to examine how a quantified definition of informativeness influences choice behavior. This was done by designing a relatively novel computerized learning and decision-making task, so called a *multi-attribute bandit task*. Bandit tasks are a popular paradigm in the cognitive neuroscience literature because it allows tight control of the parameters influence preference. One reason for designing a task in-house was, first, to engage human subjects in the task by forcing them to keep track of distinct attributes and integrate their values to successfully maximize gains. The second reason was to examine whether and the extent that quantitative informativeness was strategically utilized during the task and its effect on human choice behavior. In short, a computerized bandit paradigm was used to tease-out this use of value and information.

It should be noted that no additional estimates of attention or neural activity were collected besides the responses and response times in the behavioral task. However, this should not be interpreted as a weakness of the study, nor as if an essential research component is missing. Complex cognitive process such as information seeking require constrained tests and analyses to reverse engineer how it works, namely, to discover the questions or computational problems the brain must solve to implement cognition and behavior (Kriegeskorte & Douglas, 2018; Marr, 1982). Further, it can be argued that the development of constrained behavioral paradigms contributes more to our understanding of the brain and mind than invasive techniques (Niv, Forthcoming). More specifically, Niv argues that invasive methods, while expensive, generally tell us where in the brain neuron activity corresponds with behavior, but not what those regions are doing – behavioral paradigms allow researchers to investigate the latter question.

2.2 Methods

The bandit task is a computerized environment used in the cognitive neuroscience literature wherein subjects are asked to deliberate between distinct options that lead to rewarding or aversive outcomes (**Figure 7a**; O'Reilly & Mars, 2015). In learning versions of the task, outcomes are the only information subjects receive about their choices. From these outcomes, subjects must infer the environmental statistics so they can maximize their final reward (e.g., Behrens et al., 2007; Gershman & Tzovaras, 2018; Muller et al., 2019; Niv et al., 2015; Rouault et al., 2019). To pose such an environment on subjects in the present thesis, a multi-attribute one-armed bandit was designed. Instead of learning singular option-outcome associations, subjects needed learn the about two attributes, and then integrate this knowledge to make accurate option comparisons and maximize rewards (**Figure 7b**). The task was designed to be an online experiment, prepared using the JavaScript based library, jsPsych (de Leeuw, 2015).

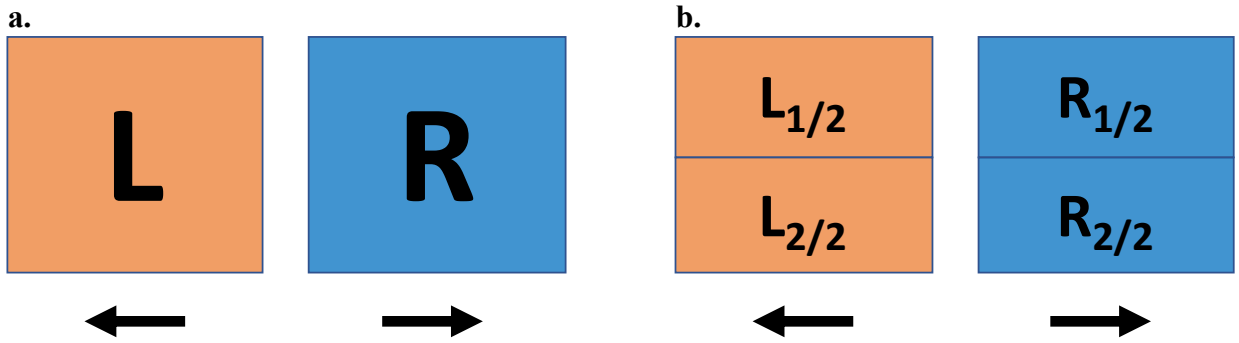


Figure 6. *Typical designs of the bandit task.* ***(a)*** *A traditional bandit task with two discrete options, where subjects can choose either the left option, L, or the right option, R. This is a typical configuration, and it allows researchers to control the parameters that influence preference (e.g., Behrens et al., 2007; Gershman & Tsovaraz, 2018).* ***(b)*** *A multi-attribute bandit task wherein options are composed of two attributes each – to choose the high reward option, both the association between each attribute and its value must be learned.*

2.2.1. Human subjects

Healthy adults were emailed an invitation to play a web-based psychology game for monetary reward. Emails were obtained by solicitation or by reference using a recruitment flyer

shared via email or the WhatsApp mobile app (**Appendix A**). Subjects were mostly students or staff from the Hong Kong Polytechnic University with a few referred to the study by a friend or family member at this the university – the subject pool is therefore a convenience sample. Of those invited, 42 subjects completed the game. Three subjects were excluded due to poor performance (subjects chose the high value option less than 60% of the time). The remaining 39 subjects (16 female) were on average 23.4 years old ranging between 18-30 years. No further demographic information was collected. This sample size was deemed appropriate because similar studies on the explore-exploit dilemma find significant with under 40 subjects (e.g., Behrens et al., 2007: 18 subjects; Wilson et al., 2014: 31 subjects). This project was approved by the Human Subjects Ethics Sub-committee (HSESC) or its delegate of The Hong Kong Polytechnic University (reference number: HSEARS20190416034).

Exclusion Criteria

Before initial consent could be granted, subjects were informed about the exclusion criteria in their invitation email. Subjects were asked to exclude themselves if they had a history of neurological impairment, vision was not normal or could not be corrected-to-normal, were younger than 18 years, were older than 40 years, or were left-handed. The reasons for these exclusions were as follow. The present thesis was interested the learning and decision-making abilities of healthy human brains. As discussed in the literature review, these processes are distributed, hierarchical, and occur in parallel in healthy brains - disruptions to any brain region may affect behavior in subtly but significant ways. Therefore, to reduce behavioral noise present between subject variation, neurological impairment was excluded; for similar reasons, age was restricted to the interval between 18-40, during the highly probable time of neural maturation and healthiness. Although stimulus salience was not manipulated, normal or corrected-to-normal

vision was critical in this study as it was required to read instructional documents and play the computerized game without a proctor to correct display issues that may arise. Finally, left-handed subjects were excluded here to improve the homogeneity of the sample pool, as commonly practiced in the cognitive neuroscience literature. The idea that handedness predicts brain differences comes from demonstrations of neuroanatomical and functional differences between left-handed brains compared to the more ubiquitous right-handed brain (Willems et al., 2014). Perhaps future versions of the task using neuroimaging can investigate left-handed performance to identify the extent that brain laterality has on multi-attribute learning and information seeking, but these potential findings were excluded from the present thesis.

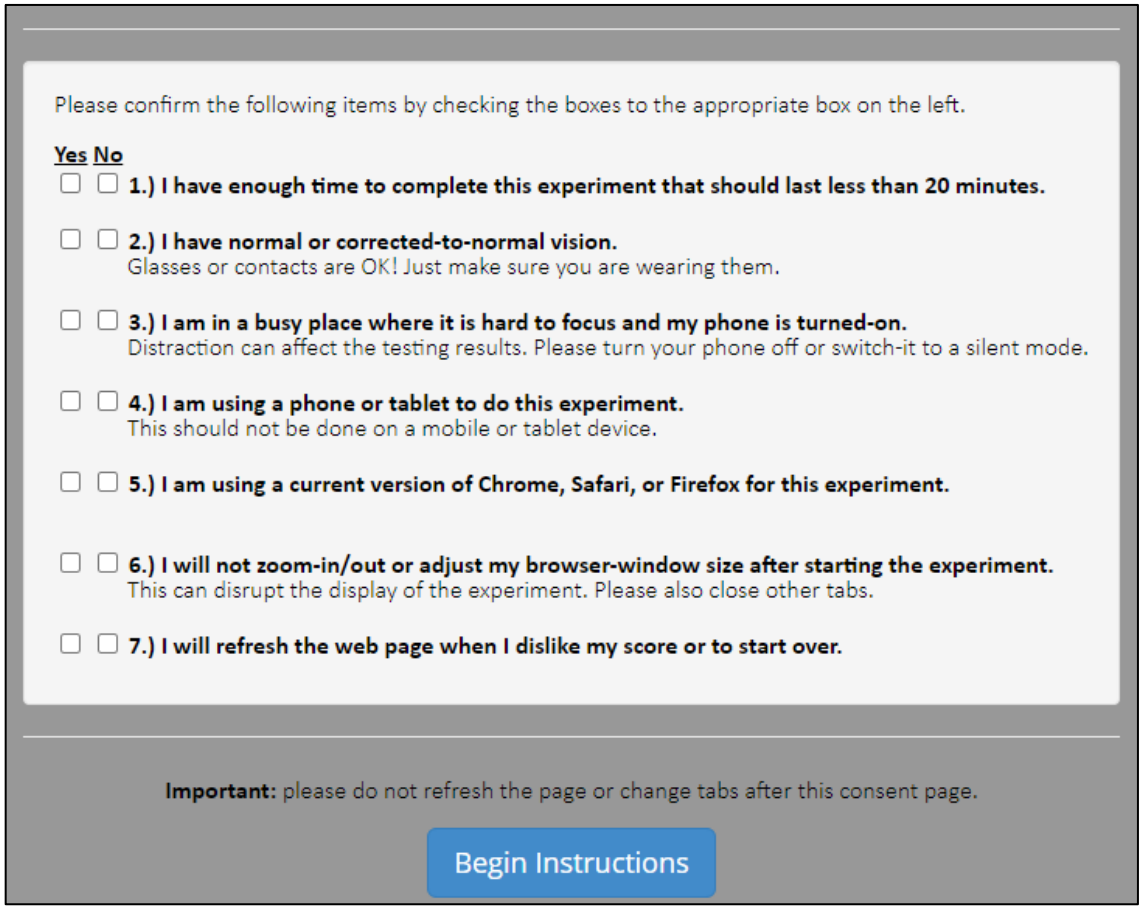


Figure 7. *Confirmation checkboxes before the start of the main task.*

2.2.2. Procedure and payment scheme

Procedure

After obtaining initial consent, subjects were emailed links to instructions and practice task. Subjects were free to study the English or Chinese versions of the instructions, then were asked to take the practice task once the instructions were understood. Practiced performance was assessed to confirm the task was understood before continuing in the experiment. Otherwise, subjects received feedback of their practice results if they failed and asked to try the practice task again. After subjects completed the practice task, they received an email with a link to the main task; at the start of the main task, subjects were asked for consent again (**Appendix C**). They could not start the main task unless they gave consent and confirmed they had time to complete the task and other requirements by reporting on checkboxes (**Figure 7**). As in the practice game, subjects returned their main task results via email – this was necessary because of problems with the server that hosted the game files prohibited the data files to be saved remotely. These data files needed to be saved locally on the subjects' computers. All results were analyzed using MATLAB (Mathworks). Subjects received their feedback about reward size via email along with a request for their "PayMe from HSBC" account information (a popular Hong Kong money exchange mobile app) to pay subjects their monetary reward. This payment was used to incentivize effort in the task and its amount was based on trial-to-trial performance in the task.

Incentivizing Performance

Subjects who completed the task automatically earned a base reward of 50HKD (6.45USD). All 42 subjects that completed the game were entitled to this base reward. As this behavioral experiment was entirely remote, instructions and the task could not be proctored. A guaranteed base reward served to incentivized subjects to independently study the instructional

materials so they can reach the end of the experiment. In addition to this completion reward, subjects were offered an additional 150HKD (19.35USD) depending on their performance in the task as determined by two criteria: 1) choice accuracy rate, defined as the percentage of trials where subjects chose the high valued option; 2) survey accuracy rate, defined as the accuracy of survey trials interleaved in the tasks, wherein subjects were asked to report running current estimate of an unknown variable in the task, described further below; and 3) the percentage of trials with a proper response, as it was possible to miss trials when a subject was too slow or not paying attention. The results for each criterion were emailed to subjects after they completed the game. In short, the final reward was determined as follows:

$$Final\ Reward = \$50 + [\$150 * (accuracy - (\text{missed trials} * \$10))]$$

This payment scheme was described before the main task. The average final reward was $133.62 and ranged between $50-$192.50. Note, $50 was the smallest amount that could be received. Subjects were warned they would only receive the $50 base reward if choice accuracy was too low. Only 1 of 42 subjects refused to receive payment. Payment was done remotely using the mobile phone app "PayMe from HSBC" as this app was widely used in Hong Kong to exchange money. Incentive schemes such as this are common in decision neuroscience experiments that aim to incentivize effortful value-based decision making and learning, improving the generalizability of results outside of the lab to real-world behavior (Chau et al., 2014; Fujiwara et al., 2018; Gluth et al., 2018; Hunt et al., 2014; Juechems et al., 2019; Nassar et al., 2010; Polanía et al., 2019; Stojić et al., 2020; Ting et al., 2015).

2.2.3. Behavioral task

Note that several terms are used here to describe the set of parameters that were manipulated in the behavioral task. To communicate the methods in the analysis clearly, these parameters are

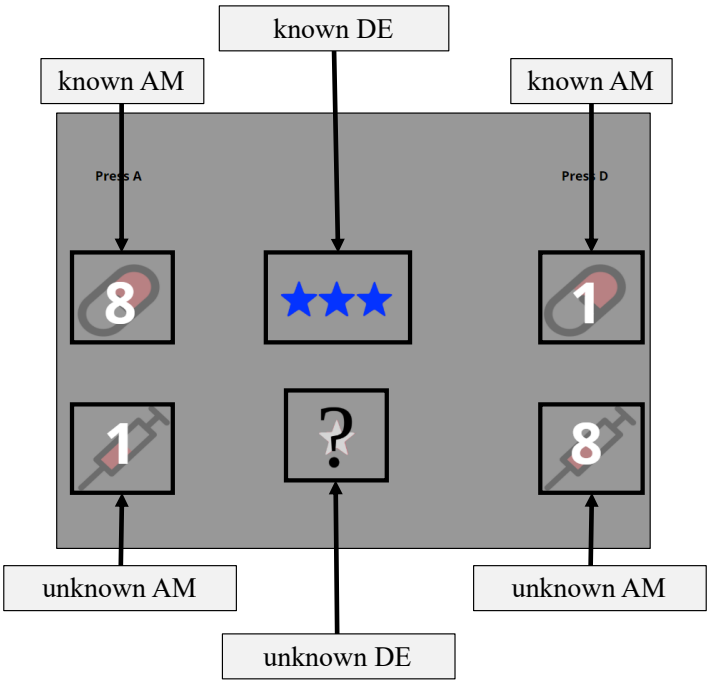
referred to using acronyms collated in **Figure 8** and are redundantly defined in the methods to avoid any confusion. To facilitate performance in the task, subjects were given a cover story where they role played as doctors treating patients with an unknown virus. Each treatment combined two drugs: a pill and syringe. On each decision trials, subjects were asked to choose one of two available treatments that combined varying amounts of pills and syringes, i.e., **attribute magnitudes** (**AMs**) refer to drug dose: the number of pill or number of syringes that make up a treatment, with doses ranging from 0 to 9. AMs were indicated by a number overlapping the corresponding drug symbol for pill or syringe (**Figure 9a**) Subjects were told the **treatment value (TV)** of each treatment was equal to the AMs weighted by **drug effectiveness (DE)**: each drug had corresponding absorbance, which was defined as the drug dose that has an effect. DE could take a value from -90% and +90% and was represented by blue or red stars for the **known DE**, or a question mark cue for the **unknown DE** (**Figure 8**; **Figure 9b**). Subjects were instructed to choose treatments using the keyboard, pressing the 'a' key to select the left treatment and 'd' key to select the right treatment. Treatment values were computed using the following equation:

$$TV = [known\ AM * known\ DE] + [unknown\ AM * unknown\ DE] \quad \text{[Equation 10]}$$

Note that the equation resembles the aforementioned *attribute integration model of subjective value computation* (Rangel & Clithero, 2014), **Equation 9** is presented again for convenience:

$$\text{Option Value} = \sum (\beta_i * \text{attribute}_i)$$

This computation determined the feedback that was displayed to subjects after choices in the decision trials (**Figure 9c**). Additionally, it provides theoretical support for the analyses, the design of the behavioral paradigm implemented in the present thesis, and discussions of the results.



Term	Meaning in the task
Attribute Magnitudes (AM)	AM is the number of doses for either the syringe or pill of a given treatment option in the behavioral task. It is the "**known AM**" if the value refers to the drug (syringe or pill) with known attribute effectiveness or it is "**unknown AM**" if the value refers to the doses of the drug with unknown effectiveness.
Attribute Effectiveness (DE)	Multiplier for each drug (distinct for the pill and syringe). A "**known DE**" indicates the value is displayed to subjects during decision trials; a "**unknown DE**" indicates the value is hidden from subjects during decision trials.
Treatment Value (TV)	Treatment Values refer to the true reward amount of an option (left or right). This value is computed by the following function for each option: (known AM*known DE) + (unknown AM*unknown DE) "**known TV**" "**unknown TV**" Here, "**known TV**" only refers to the known attribute magnitude weighted by the known drug effectiveness; "**unknown TV**" only refers to the unknown attribute magnitude weighted by the unknown drug effectiveness.

Figure 8. *Critical terms for the methods section. These terms refer to the critical parameters describing how value and informativeness were manipulated in the behavioral task.*

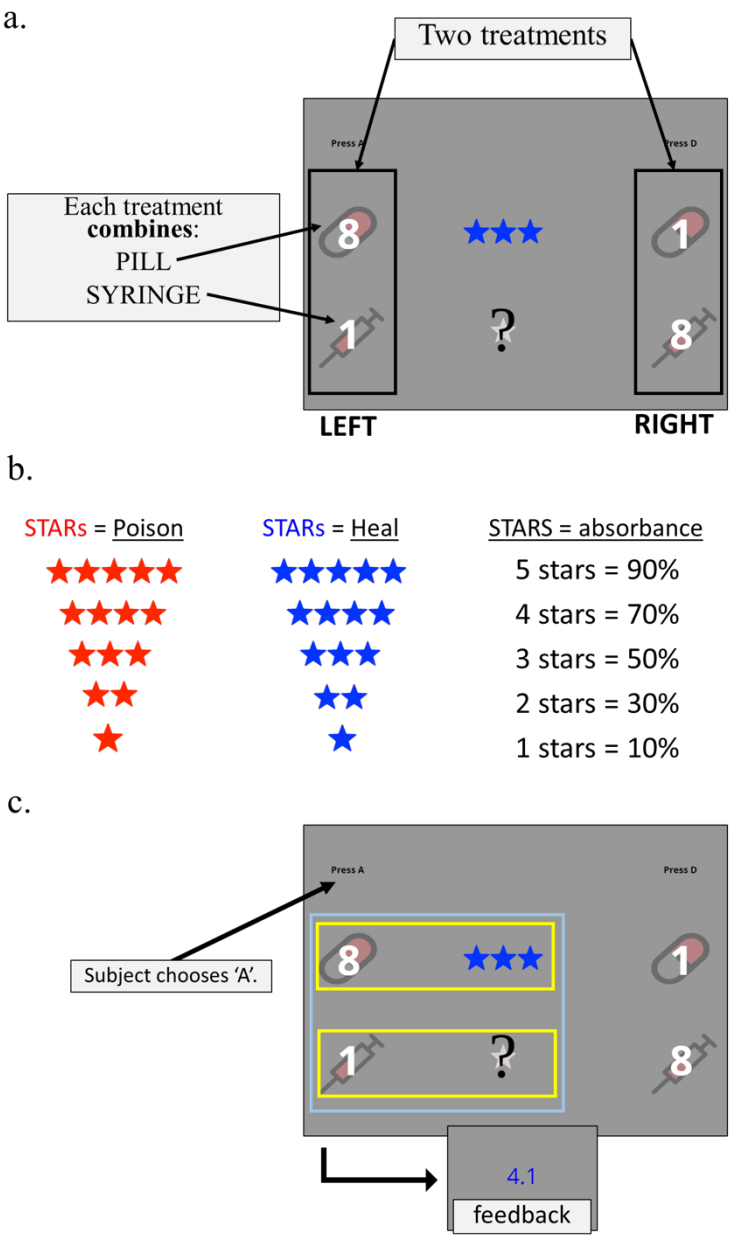


Figure 9. *Decision in the task were contextualized as treatments.* ***(a)*** *An example of a decision trial. Each option is composed of two attributes, a pill and syringe, and their respective* ***attribute magnitudes (AMs)****. In this example, the left treatment combines 8 pills and 1 syringe. Conversely, the right combines 1 pill and 8 syringes.* ***Drug effectiveness (DE)*** *is also displayed for one of the two drugs, indicated by the number of stars and color (blue for healing, red for poisonous). The other drug needs to be learned, indicated by the question-mark symbol.* ***(b)*** *Stars indicate the absorbance of the drug, its DE, which itself ranges from -90% to +90%. Negative DE symbolizes poisonous or adverse treatment and positive symbolizes healing or appetitive treatment.* ***(c)*** *Treatment values (TVs) are the sum of AMs weighted by their respective DE. AMs and their respective DE are boxed in yellow for clarity, while the sum of these gives the TV boxed in blue. In this example, the 8 pills have an absorbance of 3 stars (50%), so it has a* ***known TV*** *of 4. The 1 syringe has an absorbance of 1 star (10%), so it has an* ***unknown TV*** *of 0.1. Their sum, giving a TV of 4.1, serves as feedback for subjects.*

Throughout the task, subjects were only shown the DE of one drug, either the pill or syringe, referred to here as the **known DE**. The DE of the other drug needed to be inferred using feedback, referred to here as the **unknown DE**. Note, the DE could be negative to indicate an aversive contribution to the treatment value, TV (**Figure 9b**). Contextualized as *poisoning the patient*, subjects endeavored to maximize healing and minimize poisonous treatments, i.e., their goal was to maximize the treatment value, TV. Altogether, subjects were required to learn about the unknown DE to develop an estimate of the unknown TV in order to combine with the known

TV. Accuracy here produced proper value-based decision making. Additionally, this design allowed investigation of value based decision-making strategies while navigating a computerized multi-attribute environment.

After feedback, a small fixation button appeared at the center of the screen. Subjects were required to click this button to start the next trial (**Figure 10** shows the time flow of the task). This served two purposes: to control attentional bias that may occur if subjects endogenously due to a preference for leftward or rightward options or exogenously due to a preference to the pill or syringe stimuli (Fiedler et al., 2019; Gluth et al., 2020; Hunt et al., 2018) and to record and punish failed attention on the task, such as when subjects were away from their keyboard. As this was a remote task without a proctor, subjects were warned that missing too many trials and fixation button clicks would disqualify their participation in the experiment. No data was excluded over this scenario. Examining only the 39 subjects who completed the task with a

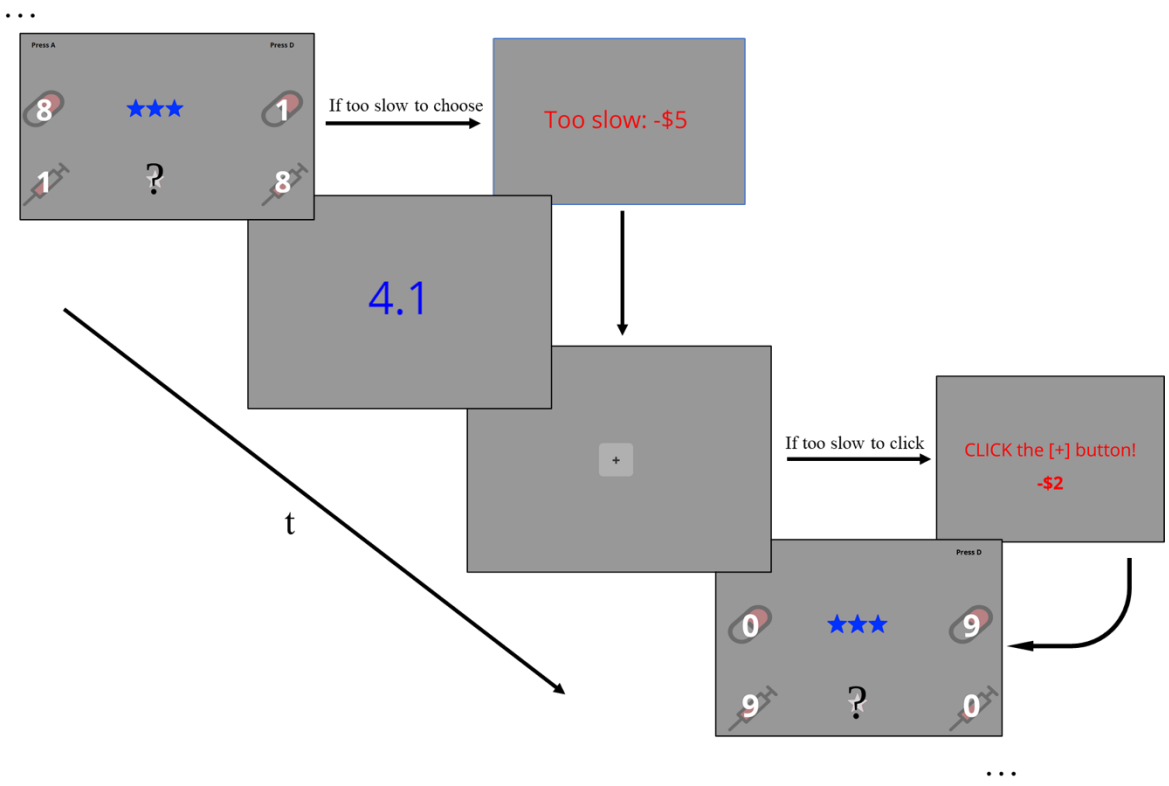


Figure 10. *The time flow of decision trials.*

choice accuracy rate above 60%, no subject missed more than 1.5% of trials, with the average rate of missed trials across subjects equaling 0.19%.

Subjects were also probed about their running estimate of the unknown DE (**Figure 11**). These survey trials were interleaved with decision trials, occurring at a pseudo-random schedule to avoid signaling any changes in DE. Subjects were instructed how to use the slider to report their running estimates. If they were certain about their estimate, they were advised to slide the white square directly over the number of stars that represents the unknown DE (**Figure 11**). However, if subjects were uncertain about their estimate in the unknown DE, they should slide the white block between these ticks, or nearer their more certain choice. For example, a subject believed +2 or +3 stars can explain their observations (**Figure 11**; see **Figure 9b** for DE values that stars represent).

Finally, subjects were instructed that the values of the known and unknown DE could take and update-to after some trials. However, they received no signal when DE updates did or might occur. The update occurred for the known and unknown DE, though not at the same time. Changes to either DE did not signal changes to the other. Subjects were notified about this update process, but not notified about the schedule or occurrence of updates. The purpose for these updates without warning was to promote attention on both attributes in the task and to examine trial-and-error learning, which makes information seeking behaviors explicit (**Chapter 1.5. – "Informativeness"**). Additionally, the location of the unknown DE swapped after a block, e.g., if subjects started the task with the unknown DE on top in the first block, the unknown DE was at the bottom in the next block (**Figure 8**). The initial location of the unknown DE was randomized.

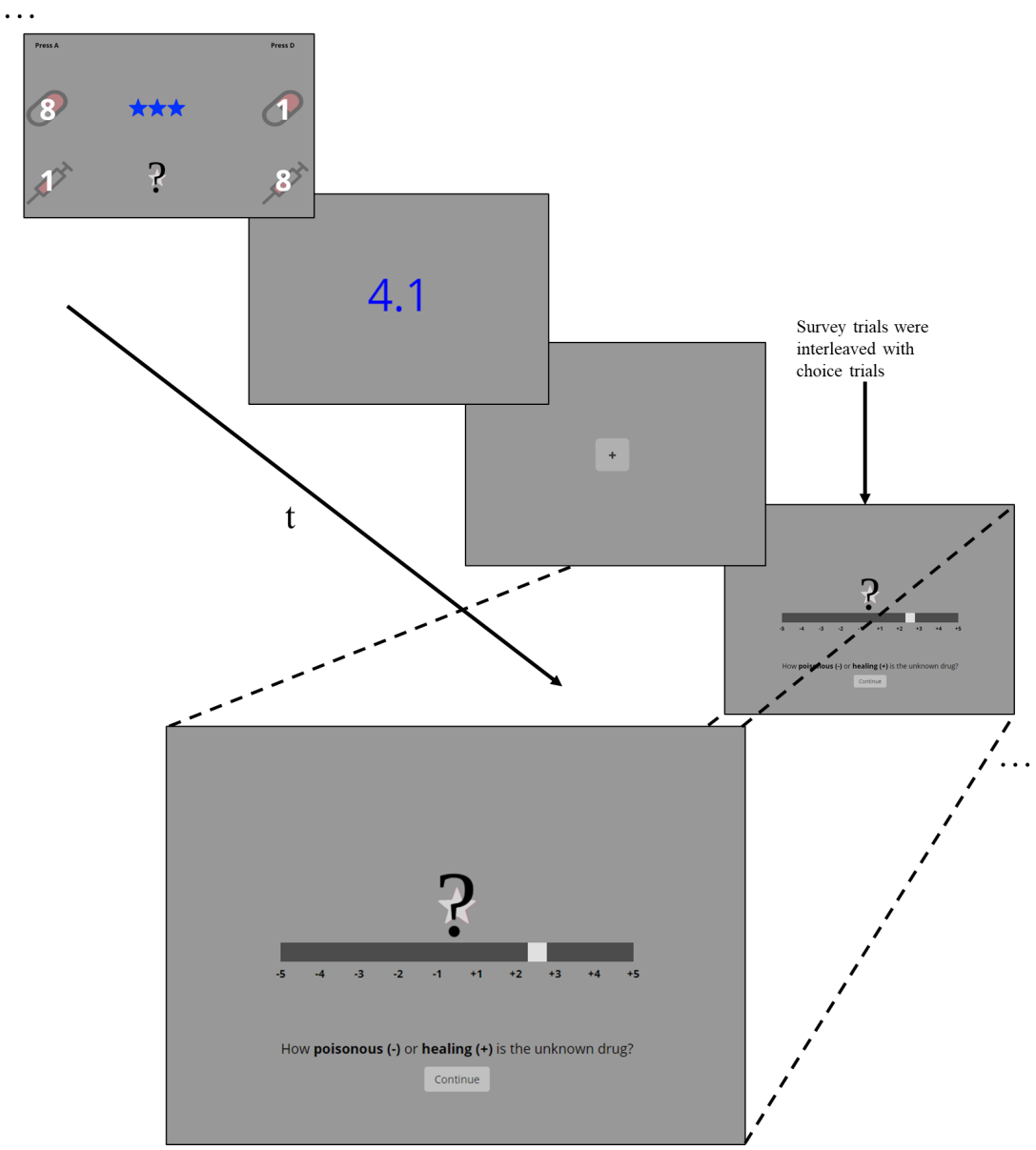


Figure 11. *Survey trials were pseudo-randomly interleaved with decision trials in the task. Pseudo-random, because these were interleaved in a manner that could not signal a change in unknown DE.*

Specifications

The practice task was composed of 30 decision trials and 4 survey trials; its schedule was the same for all subjects (shown in **Figure 12**). The purpose of the practice task was to confirm the rules of the task were understood; to ensure subjects could run the experiment on their computer without any technical issues; and confirm data could obtained.

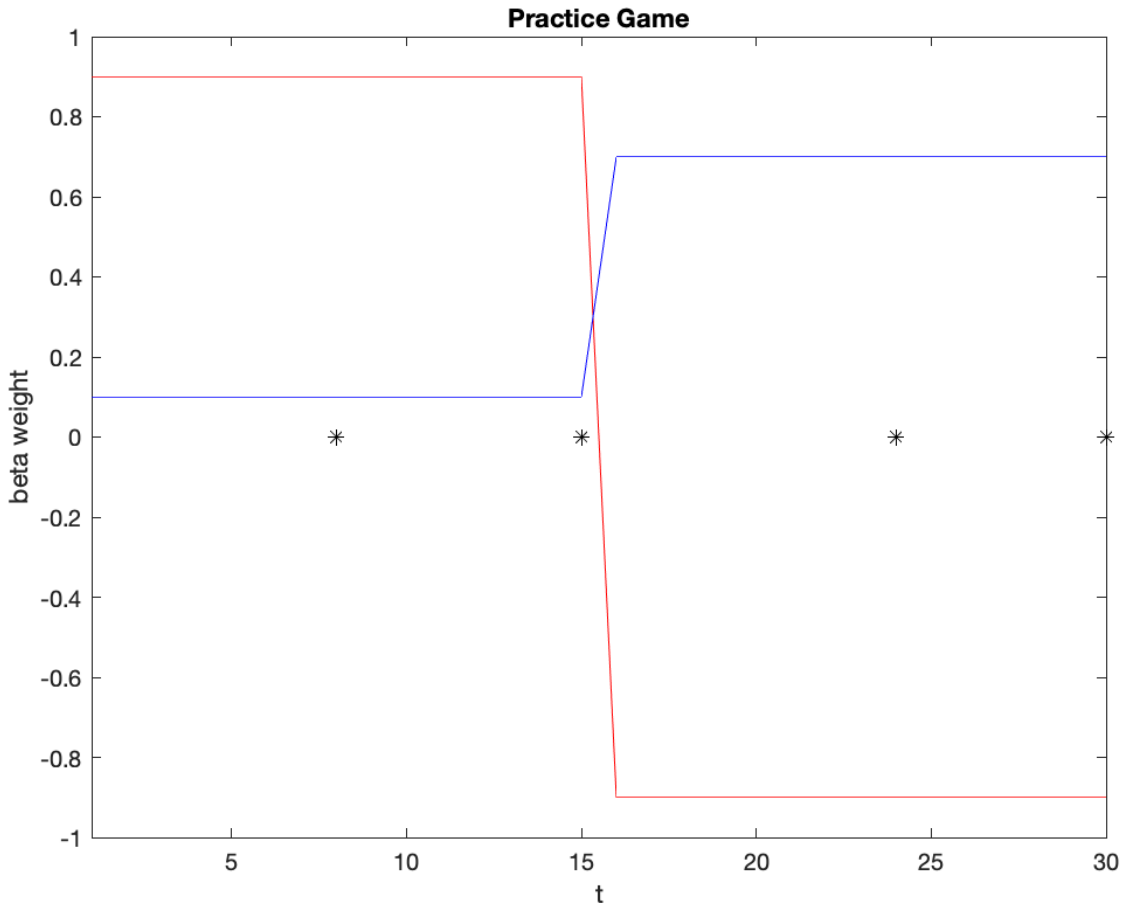


Figure 12. *The practice task schedule of decision and survey trials. The horizontal axis indicates the trial number, the vertical access represents the DE value of the known (blue line) and unknown (red line) drugs. Asterisks indicate survey trials. Survey trials probed subjects about their running estimate of the unknown DE (red line). These survey trials occurred on trials 8, 15, 24, and 30 for all subjects.*

Subjects began the main task after practice results were satisfactory. The main task was composed of 200 decision trials with 20 interleaved survey trials in one total session. One possible schedule is shown in **Figure 13** (the rest are shown in **Appendix G**) Subjects completed 2 blocks of 100 trials. Each block was characterized by whether the unknown DE was on the top (pill effectiveness) or at the bottom (syringe effectiveness) during decision trials (**Figure 10**). The subjects received a 1-minute rest period between blocks. Within a block, both the known and unknown DE updated every 20 choice trials. This update was offset between the DEs, i.e., the first DE updated on decision trial 10 and then every 20 trials thereafter; meanwhile, the other DE

first updated on trial 20 and then every 20 trials thereafter. DE updates were indicated by a change in the number of stars and color on the same row as the corresponding drug and were constrained to a specified set of stars (**Figure 9b**). Subjects were not told nor signaled when DE updates would occur in the task. The schedule for survey trials was the same across subjects regardless of assigned decision trial schedule. Survey trials were interleaved with decision trials and occurred twice within the duration of each unknown DE: once during the first ten trials after the unknown DE updated and again during the last 10 trials after an unknown DE update. The survey trial involved moving a white-box slider over a 100 ticked black bar with 10 possible DEs displayed; one for each star level (**Figure 9b**).The survey trials served as an independent measure for whether the models made accurate estimates of the unknown DE and used to confirm that the observed human behavior was due to understanding the rules of the task. Lastly, subjects were aware about time pressure during the survey t rials but were not aware of its exact duration of 15 seconds. Fixation buttons needed to be left-clicked within 5 seconds or subjects would receive feedback about their slow response and the amount of money lost before starting the next decision trial. Decision trials also had time pressure. Subjects pressed ‘a’ key or ‘d’ key to select the left or right treatment, respectively, within 15 seconds or be penalized $10. If subjects missed the decision trial, they received feedback that they were too slow, and the amount of money lost (**Figure 10**). There was no time pressure for the web pages preceding the main task, e.g., subjects were free to read the consent page (**Appendix C**) and take as long as needed to fulfill the checkboxes confirming the subject is under appropriate conditions to start the main task (**Figure 7**).

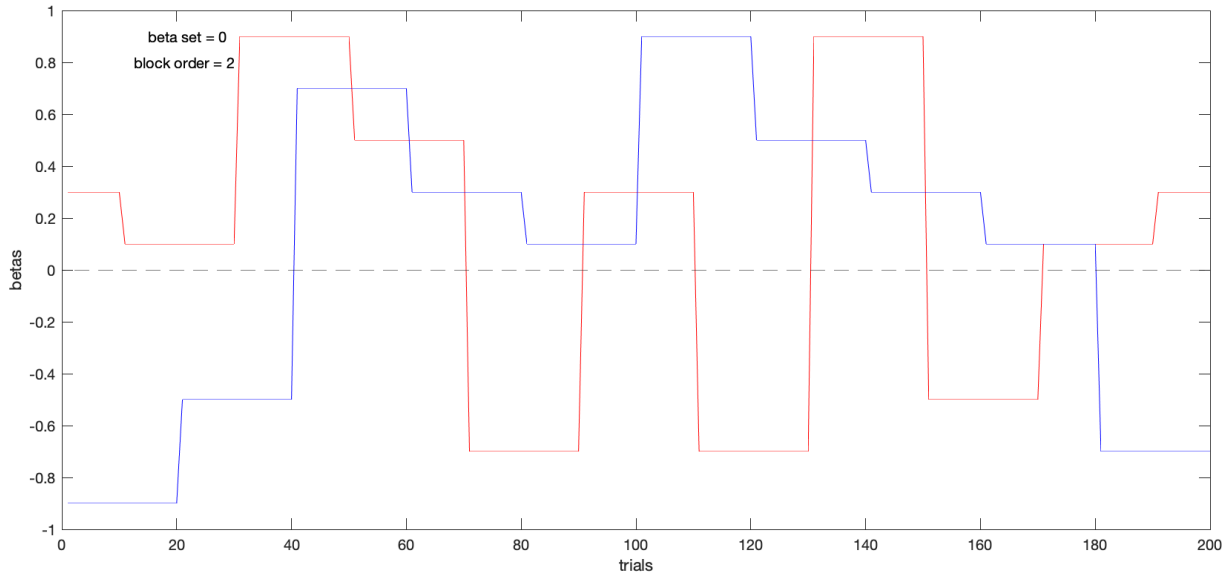


Figure 13. *One possible schedule of the DE ("betas") of the main task over 200 trials. Subjects were randomly assigned a schedule when the link to the main task was clicked. The red line indicates the schedule of the unknown DE and the blue line indicates the known DE. All the DE schedules that could be assigned are presented in* ***Appendix G****.*

According to the cover story that was presented to subjects, the main task was a series of treatments, where each treatment was a consequence of the choices made between two treatments in the decision trials. **Figure 14a** displays an example of an easy decision trial, which assumes that subjects learned the unknown DE (star values in **Figure 14a** and **Figure 14b** are displayed for didactic purposes). Easy decision trials occurred when the TV difference between options was large, as when the known and unknown DE differences was high. Conversely, decision trials were hard when the known and unknown were similar as in **Figure 14b**. **Figure 14c** depicts a more typical scenario. In addition to manipulating expected reward (the TV), option informativeness was also manipulated. Informativeness was manipulated by varying the attribute magnitudes (AMs), i.e., the number of doses of each drug composing a treatment. Informativeness was assumed to depend on the AMs, as these and feedback were the only known information for subjects to use to facilitate value-based decision making. Here, it was speculated

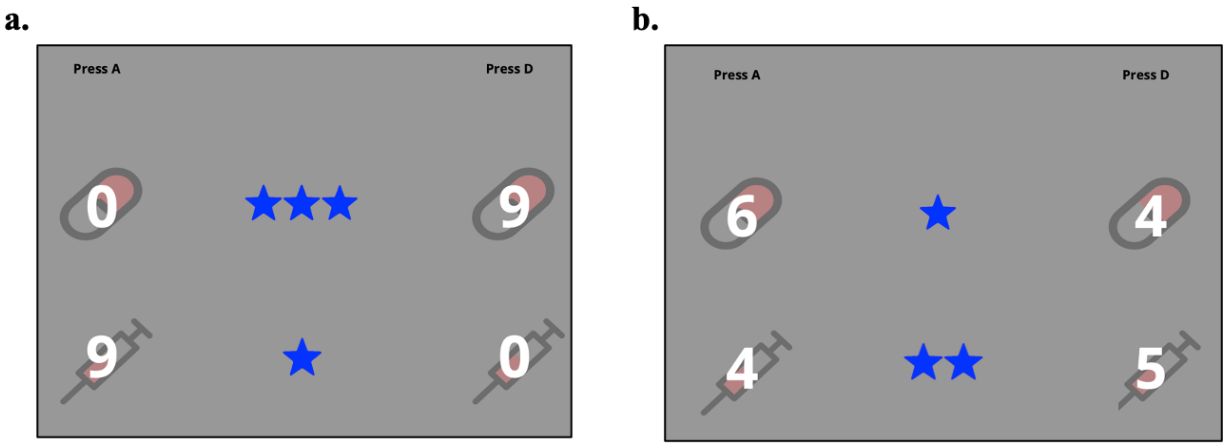


Figure 14. *Examples of easy and hard value-based decision making.* ***(a).*** *An example of an easy choice trial. Both DEs are shown here for clarity, so that this example assumes the subject knows the value of the unknown DE. In this example, the pill DE (3 stars; 50% of the AM heals) is greater than the syringe DE (1 star; 10% of the AM heals). Since the right option, selected by pressing the 'd' key, includes more pills, its treatment value is greater, i.e., 4.5 (right, 'd' option) compared to 0.9 (left, 'a' option). These treatment values were the only feedback subjects received regarding their choices in the game. Note that this decision is easier than the decision trial shown in (b).* ***(b)*** *An example of a harder decision trial that (a), where both DEs are shown here for clarity. Here, it is harder to determine which option has a higher TV. In this case, the right option returns 4.9 treatment value whereas the left option returns 4.2. It should be harder to choose the high TV when its similar to the low TV.*

that a higher, known AM would be more useful when compared to feedback, e.g., if the known drug dose was large, but feedback was small (the rational is discussed below). This was formulated as follows:

$$informativeness = \frac{known\ AM}{known\ AM + unknow\ AM} \qquad \text{[Equation 11]}$$

Information should be the most appetitive soon after an update to the unknown DE is detected, as seen in the literature on the explore-exploit dilemma (Wilson et al., 2020). Notably, this estimate of informativeness is more quantifiable that the categorical measures in the literature (e.g., Trudel et al., 2020; Warren et al., 2017; Wilson et al., 2014; Zajkowski et al., 2017). As discussed in **Chapter 1.5 – "Informativeness"**, informativeness is an option attribute whose outcome results in the reduction of uncertainty. The rational of **Equation 11** is represented in **Figure 15**, which shows an example decision trial. In a scenario where the subject is uncertain about the

unknown DE, such as when a change has been detected, subjects should make decisions that makes facilitates learning – in this case, they should make decisions that allows them to maximize how much of the expected feedback they understand. This is accomplished by choosing the treatment with the highest known AM (this is left option in the decision trial in **Figure 15**) and makes examining the contribution of the unknown TV easier to interpret. For example, if the known AM is maximized and feedback is observed to be mostly from the known TV, then the known DE is likely greater than the unknown DE; conversely, if the known AM is maximized but most of feedback is not explained by the known TV, then it is likely that the unknown DE is greater than the known DE. The consequence of this decision strategy is an accelerated reduction of uncertainty following feedback; this behavior is an example of the definition of informativeness: the reduction of uncertainty, albeit an indirect measure (**Chapter 1.5. – Informativeness**).

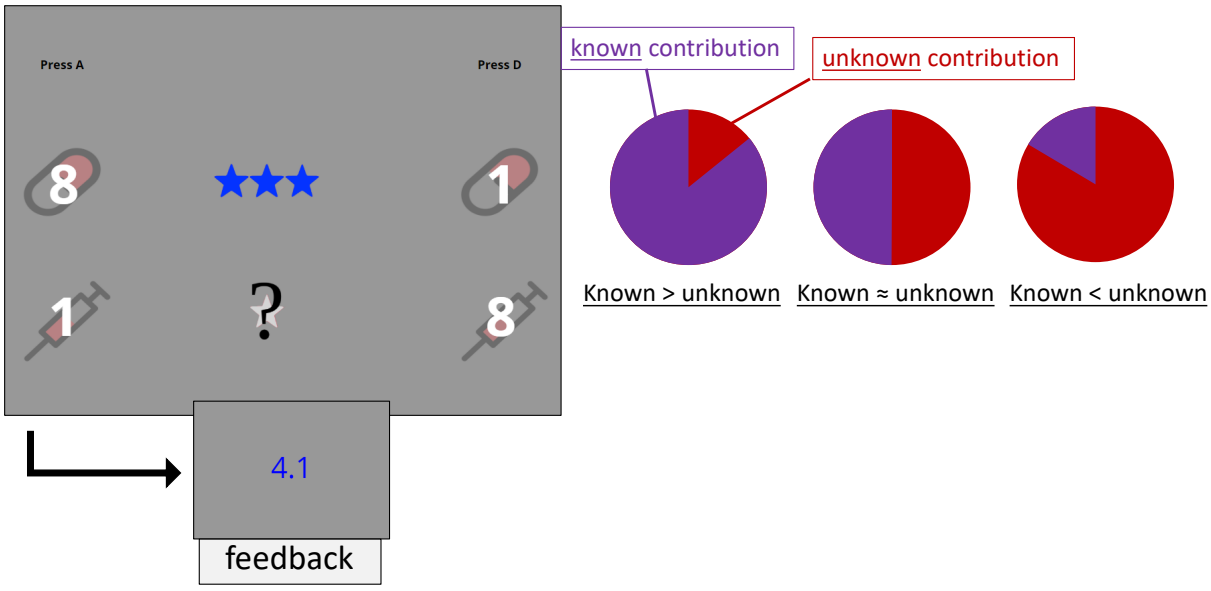


Figure 15. *Rationale for the quantitative measure of informativeness in this thesis. Informativeness should be the most appealing after an unknown DE has been detected. During this period of high uncertainty, decisions should focus on facilitating learning from the outcome.*

2.3. Results

A reminder that the critical terms describing the manipulated variables in this analysis are described in **Figure 8**.

In order to control the difficulty of the main task, DE schedules were constrained to eight possible sets (**Appendix G**). In a pilot run with the same task but with AMs and DEs generated randomly, half of eleven subjects finished with choice accuracy rates below 60% – which would have disqualified them from further analysis. The average choice accuracy was 67%±5% in this pilot sample. These preliminary results suggested that the task was too difficult. To make the task easier and retain more subject data, AM and DE schedules were generated pseudorandomly so that AMs of opposing treatments were more mirrored. That is, more trials contained AMs in which one attribute had the same magnitude on the opposite attribute in the alternate treatment, allowing for treatments to be compared more easily (e.g., **Figure 14a**). However, this increased the potential for problematic multicollinearity. Multicollinearity needed to be reduced to avoid confounds from unintended patterns between the parameters (i.e., known and unknown AMs, DEs, and TVs). **Figure 16a** presents a data array with correlation coefficients represented by scaled color matrixes; these correlations compared TVs and AMs across the eight possible schedules. **Appendix D** displays the corresponding correlation coefficients and standard errors numerically. The bottom-left quartile in **Figure 16a** is relevant in this discussion (with the bottom-left quartile in **Appendix D** serving as a numerical representation). Therein, each AM was compared with that same AM weighted by its respective DE (i.e., the known and unknown TVs, the amount of treatment contributed by the drug), showing that correlation coefficients are all below 0.2 for between attribute magnitudes (pill or syringe) and treatment value schedules. To ensure that multicollinearity was not a significant confound when using these schedules in a

logistic regression, **Figure 16b** shows variance inflation factors (VIF). In this analysis, regressors were the AMs and their respective TVs after weighting with the respective DE. VIF is measure of multicollinearity (e.g., O'Hora et al., 2015) wherein values above five indicate problematic multicollinearity between the regressors. VIF analysis produced values well below this problematic threshold (**Figure 16b**).

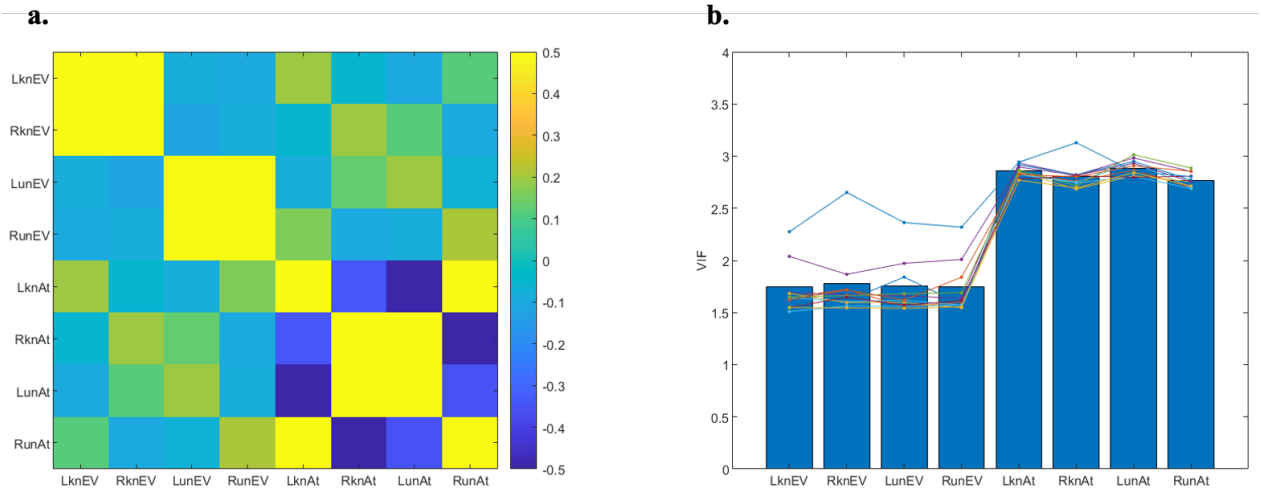


Figure 16. *Intercorrelations of TVs (AMs weighted by respective DEs) and the attribute magnitudes (unweighted AMs) for known and unknown treatment options.* ***(a)*** *The color matrix is easier to understand as a set of quadrants. In the top-left quadrant, correlation coefficients between known and unknown TVs are compared between known and unknown AMs for the left and right option (giving 4x4 matrix). Naturally, correlations are high (near +0.5) when comparing AMs weighted by the same DE, i.e., weighted by the known or unknown DE. In these cases, DE is shared for the left and right treatments. Comparing TVs across DEs, i.e., comparing unknown with known weighted AMs by DEs, correlation coefficients are closer to 0 (**Appendix D**). The top right and bottom left quadrants are the same analysis. Here, weighted AMs with unweighted AMs are compared. The correlation coefficients in this comparison are at or below 0.2, a good indication of decorrelation. The bottom-right quadrant compares AMs. The mirror-like constrained introduced to make the task easier is evident in this quadrant (discussed in the main text). When a known AM is small, it's counterpart (left or right) is small and conversely, when the unknown AM is high, its counterpart is high. Abbreviations (note that treatment value, TV, and expected value, EV, are analogous terms in this figure): Left and known expected value, LknEV; right and known expected value, RknEV; left and unknown expected value, LunEV; right and unknown expected value, RunEV; left and known attribute, LknAt; right and known attribute, RknAt; left and unknown attribute, LunAt; right and unknown attribute, RunAt.* ***(b).*** *The variance inflation factor (VIF) is used to identify unintended correlations between independent variables of a regression. Specifications were constrained before data collection using VIF, confirming values were all below 5, suggesting that multicollinearity is not problematic in the following analyses.*

On average, subjects chose the high valued option 82.9% ± 1.30% (**Appendix E** displays behavioral responses and schedules for each subject). During the task, subjects learned the values

of the unknown DE through trial-and-error. To examine whether learning occurred, trials were binned by five-trials after an unknown DE updated. The duration of unknown DEs was twenty trials, producing four time-dependent bins of five trials for the following analyses: specifically, the trial bins combined the first five trials, the second five, the third five, and the fourth five trials of the unknown DE duration (**Figure 17a**). An ANOVA analysis comparing these four bins revealed a significant effect of the trial time and choice accuracy ($F(3, 152) = 6.9$, $p = 0.002$). A Tukey-Kramer post-hoc analysis further suggested that performance in the first trial bin was significantly lower than performance in the subsequent bins. Additionally, the bins after the first five were not significantly different from each other, suggesting asymptotic performance after a learning phase (**Figure 17b**).

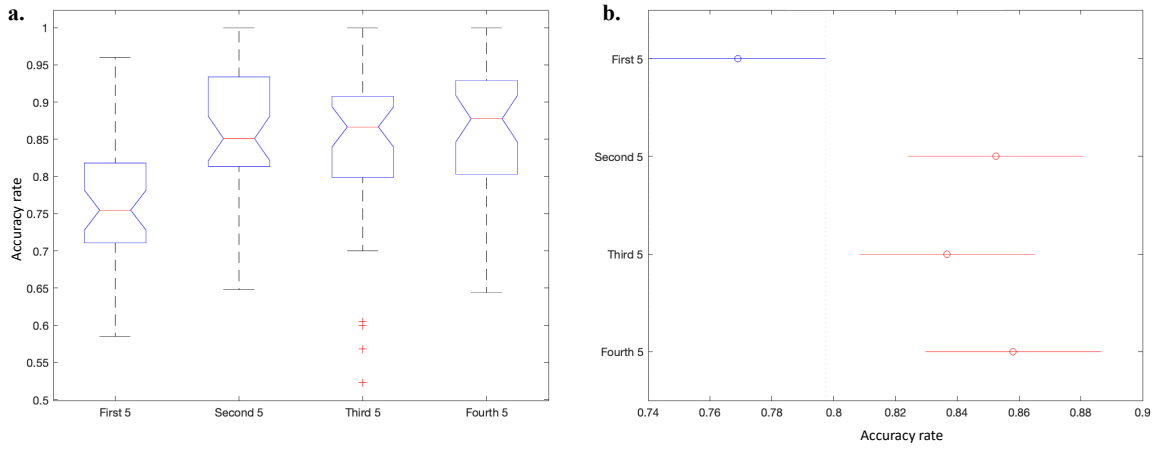


Figure 17. *Behavioral task choice accuracy analyses.* ***(a)*** *Boxplot and whisker plots demonstrating the spread of subject accuracy rates as binned by 5-trials from the onset of a new unknown DE. Red markers indicate the median and box edges indicate the first (q_1) and third (q_3) quartiles of the sample of subject accuracy means in the analysis. Values beyond the whiskers (greater than $q_3 + 1.5 \times (q_3 - q_1)$ or smaller than $q_1 - 1.5 \times (q_3 - q_1)$ are outliers and indicated by crosses. ANOVA results found a significant main effect of trial bin and accuracy ($F(3, 152) = 6.9$, $p = 0.002$; also indicated by the low overlap of the spread in accuracy in the "First 5" compared to the subsequent trial bins).* ***(b)****. Tukey-Kramer post-hoc analysis comparing accuracy rates by bin. Center circle indicate accuracy means; horizontal lines are the 95% confidence intervals. This post-hoc analysis revealed that the "First 5" trial bin had a significantly lower mean than the subsequent bins. Negative mean differences indicate an improvement in performance: First 5 - Second 5 mean difference = -0.083; $p = 0.0009$; First 5 – Third 5 mean difference = -0.068, $p = 0.01$; First 5 – Fourth 5 mean difference = -0.089, $p = 0.0003$. Further, the differences between bin after First 5 were not significant (Second 5 – Third 5 mean difference = -0.016, $p = 0.89$; Second 5 – Fourth 5 mean difference = -0.001, $p = 0.99$; Third 5 – Fourth 5 mean difference = -0.021, $p = 0.77$).*

Subjects were asked twice to give running estimates for each unknown DE: once during the first ten trials of the duration of the unknown DE, and again during the last ten trials. Subjects were also asked to report their uncertainty: the number of stars nearest to their preferred slider response. These survey trials were considered accurate it they were equal or nearest to the true DE (**Figure 11**). Using this definition of survey accuracy, subjects had an average survey accuracy of 58.9% ± 4.57%. An ANOVA analysis comparing survey accuracy rates interleaved in the first ten trials (accuracy: 52.1% ± 4.63%) with those interleaved in the last ten choice trials (accuracy: 67.1% ± 4.85%) showed a significant difference ($F(1, 76) = 5.1$, $p = 0.02$). Providing further support for the occurrence of learning in the task.

Though treatment values and their drug dose attributes were decorrelated (**Figure 16**), it was still conceivable that subjects chose according to drug dose magnitude. Further, as learning was observed, it was critical to investigate the trials where learning may have occurred; namely, whether information seeking was evidence in the "First 5" bin (**Figure 17**). To conduct this examination, a multiple logistic regression was conducted with weighted AMs (i.e., known and unknown TVs) and AMs as regressors (terms can be reviewed in **Figure 8**). The dependent variable was a binary variable indicating whether the treatment on the right was chosen or not (in other words, whether the 'd' key was pressed or not during decision trials); trials where subjects were too slow to make decisions were omitted. Regression results demonstrate that subjects generally based decision on the known and unknown TVs and less-so on AM (**Figure 18**), suggesting learning occurred and that increasing performance in the task was due to adherence to the rules of the behavioral task. Note that the **Figure 18** regressors and dependent variable vectors are arranged in five trial bins, like the accuracy analysis described above. This made explicit the evolving preference for the right option more when right known and unknown TVs

were greater, as opposed to greater respective AMs. Conversely, subjects increasingly preferred the right option less when left known and unknown TVs, and not the respective AMs, increased in value (**Figure 18**; **Appendix F**). As discussed above, unknown DEs were typically in effect for twenty trials, producing four bins of five trials each (**Figure 18**). Put concisely, decisions corresponded with weighted AMs over time as evidenced by the observation that subjects increasingly chose the right treatment more over-time when the right drug was greater and conversely chose the right option less when the left drug greater (**Figure 18**). These effects were notably greater for bins after the first five trials, compared to the second bin and onwards, suggesting that the increase in accuracy rates after the first 5 trials was due to an increasing adherence to basing decisions on the known and unknown TVs. Additionally, those regression results showed decreased and sustained independence from AMs on decisions after the first trials. That is, subsequent bins showed small and insignificant beta coefficients for all AM regressors, known and unknown. However, the beta coefficients in the first five bin for left unweighted drug doses were significant or approached significance, indicating these attributes were initially influential during decision making. One reason for this trend may be that subjects made more random decision that incidentally correlated with AMs; or perhaps AMs were used strategically during decision making when the unknown DE had recently updated. That is, subjects may have used attribute magnitudes to tease out the value of the unknown DE when its true value was the most uncertain, e.g., in a manner that made their consequence informative. This would be expected if exploratory behavior were engaged in the task, as it would have to occur in the first 5 trials when the unknown DE had updated, and feedback no longer coincided with expected TV. This was the basis of the following analysis.

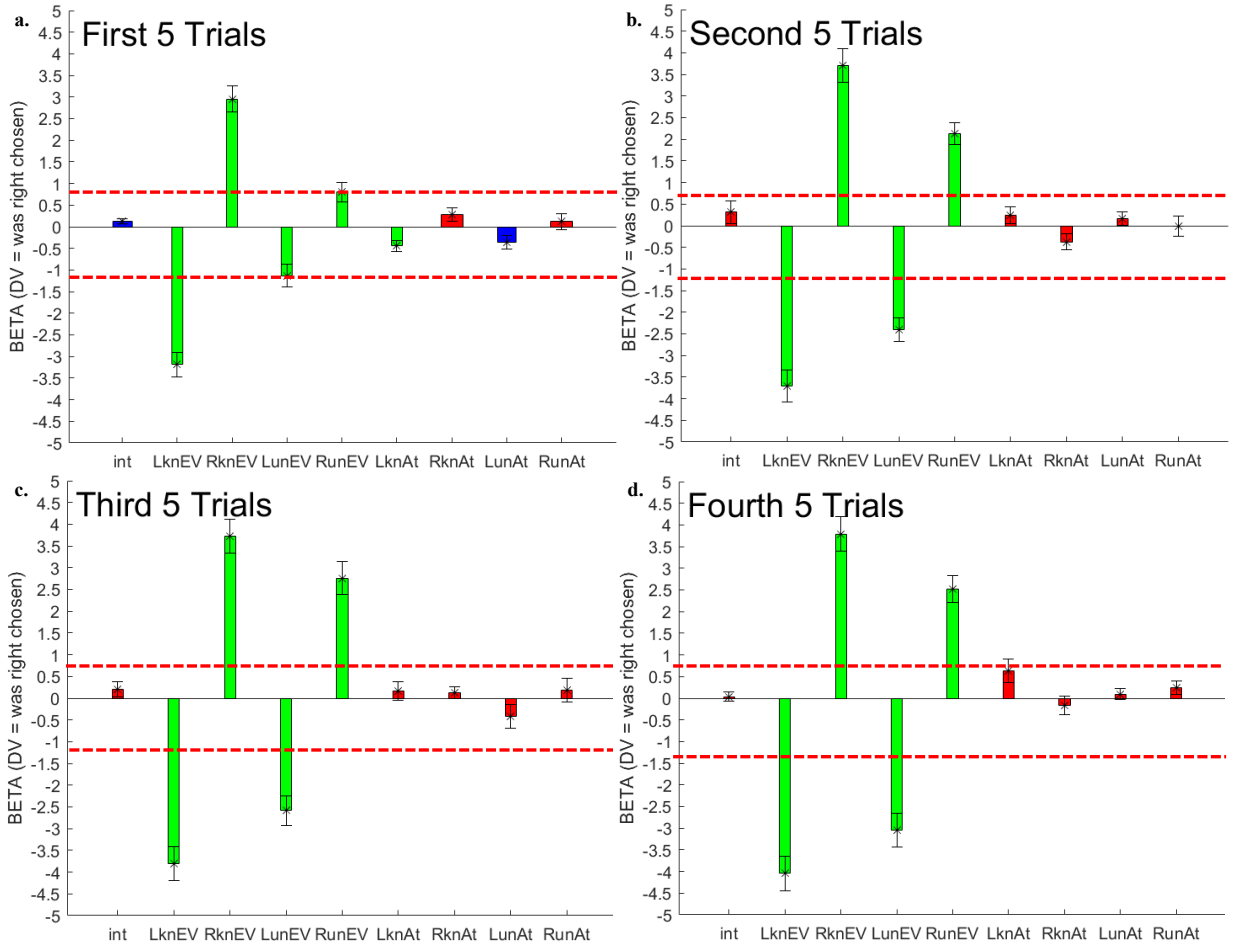


Figure 18. *Multiple logistic regression with weighted and unweighted AMs as regressors and a binary dependent variable indicating whether subjects chose the right option (equals 1) or not (equals 0). Abbreviations (note that treatment value, TV, and expected value, EV, are analogous terms in this figure): Intercept, int; Left and known expected value, LknEV; right and known expected value, RknEV; left and unknown expected value, LunEV; right and unknown expected value, RunEV; left and known attribute, LknAt; right and known attribute, RknAt; left and unknown attribute, LunAt; right and unknown attribute, RunAt. Missed trials were omitted in this analysis. Bars indicate the mean of the beta coefficient; error bars indicate SEM; more positive or negative bars indicate greater effect of the weighted or unweighted AM on choosing or avoiding the right option. Each bar plot presents results for 5 trial bins of the 20-trials an unknown DE was in effect. Green bars indicate $p < 0.05$, blue bars indicate $p < 0.10$ but $p > 0.05$, and red bars indicate $p > 0.10$. The red, horizontal, and dotted lines indicate the mean of the beta coefficients for the effect of the weighted unknown AMs during the first 5 trials. The same beta coefficients in the subsequent 5 trial bins (**b**, **c**, **d**) surpass these lines, indicating greater effect of the weighted unknown AM with subsequent trials, while the unknown DE stays the same. Notably, attribute magnitudes (the regressors labelled LknAT, RknAt, LunAT, and RunAt) do not approach significance after the first 5 trial bin, suggesting subjects did not depend on these values after this early period of an unknown DE. The effect of known weighted AMs remained high and significant, indicating subjects understood the task.* ***Appendix F*** *displays logistic regression results in further detail.*

The next analysis examined whether information seeking was conducted. Like value, the option treatment value (TV), one treatment typically had a higher treatment value than the alternate. Likewise, it is proposed here that one treatment was also typically more informative than the other. This discussion describes two definitions of informativeness – both are presented here: the first is to demonstrate evidence for why it is not a suitable definition to explain subject behavior; the second is the definition of informativeness discussed in the methods section of this chapter (**Equation 11**). To analyze treatment informativeness and tease out its consequential behavior, the first definition of informativeness was proposed as follows:

$$informativeness = \frac{unknown\ AM}{unknown\ AM + known\ AM} \qquad \text{[Equation 12]}$$

This ratio refers to the drug doses, or also the attribute magnitudes (AMs; **Figure 8**). The numerator is simply the unknown AM, divided by the denominator, which is the sum of the unknown AMs that compose a treatment. Consequently, the value of this ratio is high when the unknown AM is high. The logic for this first definition was as follows. When the unweighted AM of the unknown DE is greater than the unweighted AM of the known DE, it would be easier for the subject to learn about the unknown DE because its contribution to outcome should be larger. To take an extreme case as an example, when the unknown AM equals the maximum magnitude of 9 and the known AM equals the minimum magnitude of 0, the result of entering these values into equation 12:

$$\frac{unknown\ AM}{unknown\ AM + known\ AM} = \frac{9}{9+0} = 1$$

In this scenario, the treatment's informativeness is maximal. That is, if subjects chose an option with this informative value, then the TV outcome would be entirely the result of the relationship between the unknown AM and unknown DE. Learning is maximal in this case. In the opposite

extreme, when the unknown attribute equals 0 and the known attribute equals 9, the resulting computation is as follows:

$$\frac{unknown\ AM}{unknown\ AM + known\ AM} = \frac{0}{0+9} = 0$$

In this scenario, the treatment's informativeness is minimal. That is, if subjects chose an option with this informative value, then the TV outcome would be entirely uninformative with regards of the relationship between the unknown AM and unknown DE. Consequently, this treatment would provide no information and so learning could not occur in this case. Note that, although a treatment with these latter AMs would teach nothing about the unknown DE, it may be tempting given a high known DE. Therefore, if informativeness was irrelevant, then decisions should be entirely value-based throughout the task, and this should be observed regardless of the five-trial bin (i.e., time during the duration of an unknown DE). Conceptually, this definition of informativeness (**Equation 12**) in the task appears to be reasonable and comparable to **Equation 11**. However, they are complementary and not equal. This was easy to see in the following analysis for effects of informativeness on decision making in the task. As before, decision trials were binned by five trial for a second multiple logistic regression. Here, the four regressors included left and right option treatment values and left and right informativeness of options in each trial. As before, the dependent variable was binary and defined as whether the right option was chosen on a given trial. To be clear, the dependent variable equaled a value of one on decision trials where subjects chose the right option and equaled zero on decision trials where subjects chose the left option. The decision trials where subjects failed to make a choice, i.e., trials where the response was too slow, were excluded from the analysis (**Figure 19; Appendix F**). The results of informativeness on decision making (**Equation 12**) appeared counterintuitive to the findings of information seeking in the literature (**Chapter 1.5. – "Informativeness"**).

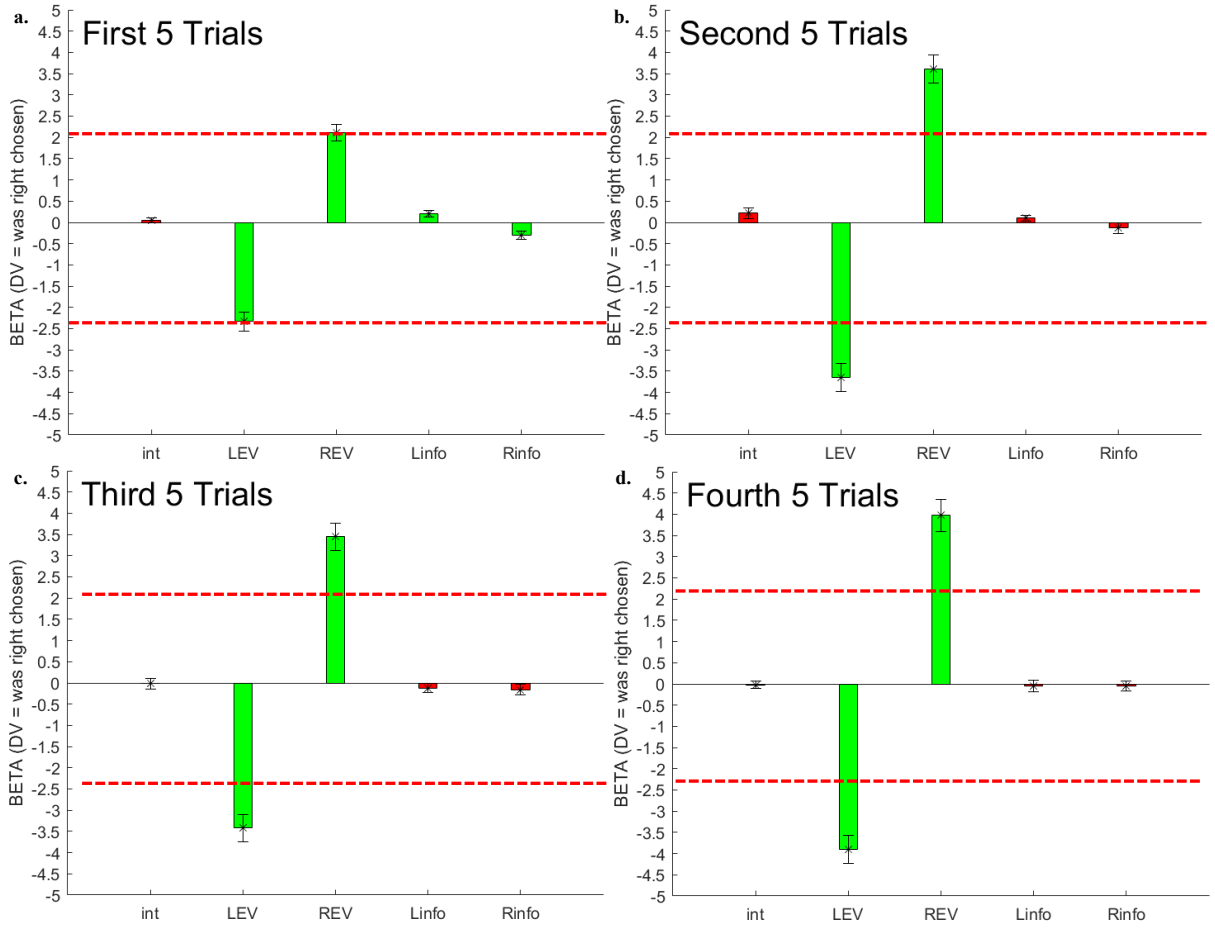


Figure 19. *Informativeness had a significant effect only on the first 5 trials, however the results appear counterintuitive. There were no further significant effects of informativeness on decision making after the first 5 trials. Multiple logistic regression with treatment values (left and right) and treatment informativeness (left and right) as regressors and a binary vector indicating whether subjects chose the right option or not. Abbreviations: Intercept, int; Left treatment value, LEV; right treatment value, REV; left treatment informativeness, Linfo; right treatment informativeness, Rinfo. The red, dotted, horizontal bars indicate the level of the mean of beta coefficients for LEV and REV in the first 5 trials* ***(a)*** *to demonstrate the progressive development of the beta coefficient in the subsequent trials* ***(b, c, d)***. *Green bars indicate $p < 0.05$; blue bars indicate $p > 0.05$ but $p < 0.10$; red bars indicate $p > 0.10$. Informativeness is defined as the unknown AM of a treatment divided by the sum of the unknown and known AMs composing the treatment option.* ***Appendix F*** *displays logistic regression results in further detail.*

Similar to **Figure 18**, the predictiveness of TVs on choosing the right treatment was negative with regards to the left TV (labeled "LEV" in **Figure 19**), such that greater left option TVs decreased the odds of choosing the right option; the converse was true for the right TV (labeled "REV" in **Figure 19**), where greater right option TVs increased the odds of choosing the right option (**Figure 19**). The effect of informativeness was significant only in the first five trials,

however, this result was counterintuitive. To elaborate these effects of informativeness on decisions (i.e., the "Linfo" and "Rinfo" regressors), when the left treatment was more informative, subjects appeared to avoided it and choose the right option. Conversely, if the right treatment was more informative, subjects avoided this highly informative right treatment. Additionally, the regressors for left and right TV with informativeness in the first five trial bin were additionally not correlated (left Pearson $r_{LEV\text{-}Linfo} = 0.10$, right Pearson $r_{REV\text{-}Rinfo} = -0.03$; subject-by-subject correlations all had p-values greater than 0.05). As discussed in the literature review (Chapter 1.5 – "Informativeness", if informativeness was a factor, then it is expected to be appetitive during exploratory phases. These phases are critical in the first bin. However, informativeness was not appetitive; it was instead aversive to value-based decision making in the task (**Figure 19**). It is clear that informativeness is realistically described by the complement of **Equation 12**. This complement is **Equation 11** that was discussed in the methods section of this chapter, repeated below for convenience:

$$informativeness = \frac{known\ AM}{known\ AM + know\ AM}$$

To be clear, the difference between these formulae is whether the numerator reflects the magnitude of the known or the unknown AM. Using the former definition of informativeness results in a rational pattern of decision making that is expected in a learning task. Specifically, the significant appetitive effects of informativeness during the first five trials are consistent with the literature on the explore-exploit dilemma (**Figure 20; Appendix Fc**), as this is the period when information seeking behavior is typically observed. To elaborate, left treatments with increasing informativeness during the first five trials significantly reduced the odds of choosing the right treatment; conversely, right treatments with increasing informativeness during the first

five trials significantly increased the odds of choosing the right treatment. Notably, these effects did not extent to subsequent bins (**Figure 20**).

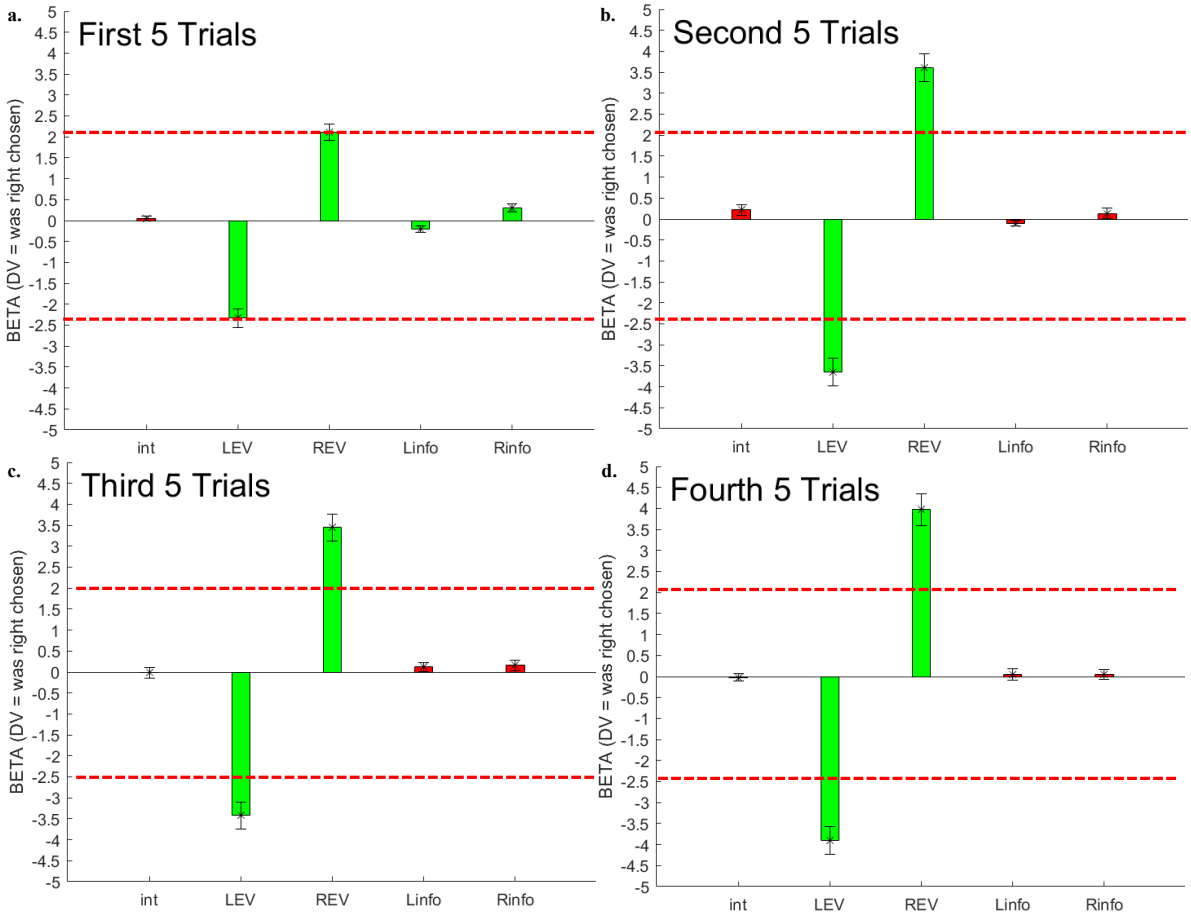


Figure 20. *Multiple logistic regression with treatment values (left and right) and treatment informativeness (left and right) as regressors and a binary dependent variable indicating whether subjects chose the right option or not. Abbreviations: Intercept, int; Left treatment value, LEV; right treatment value, REV; left treatment informativeness, Linfo; right treatment informativeness, Rinfo. The red, dotted, horizontal bars indicate the level of the mean of beta coefficients for LEV and REV in the first five trials* ***(a)*** *to demonstrate developments of the beta coefficient into the subsequent trial bins* ***(b, c, d)***. *Green bars indicate $p < 0.05$; blue bars indicate $p > 0.05$ but $p < 0.10$; red bars indicate $p > 0.10$. Informativeness is defined as the known AM of a treatment divided by the sum of known and unknown AMs composing the treatment.* ***Appendix F*** *displays logistic regression results in further detail.*

Perhaps the above reasoning for using **Equation 11** over **Equation 12** may seem like a speculative reasoning. One consequence of assuming **Equation 12** as the definition of informativeness is the case when the known AM equals zero. Here, as discussed above, informativeness would be maximal. If subjects were using the unknown AM to make informative

choices, they would exploit this case of maximum informativeness when the known AM is zero. To examine whether this occurred, an ANOVA analysis was conducted with four vectors, one for each five-trial bin that an unknown DE was in-effect as in the previous analyses, with rates based on the percentage of trials that subjects chose the maximal informative choice (**Figure 21**). There were no significant effects ($F(3, 152) = 2.51$, $p = 0.061$), suggesting that subjects neglected these maximally informative options and supporting using **Equation 11** implementation of informativeness with the known AM as the numerator for further analysis (**Chapter 3 – "Bayesian models"**).

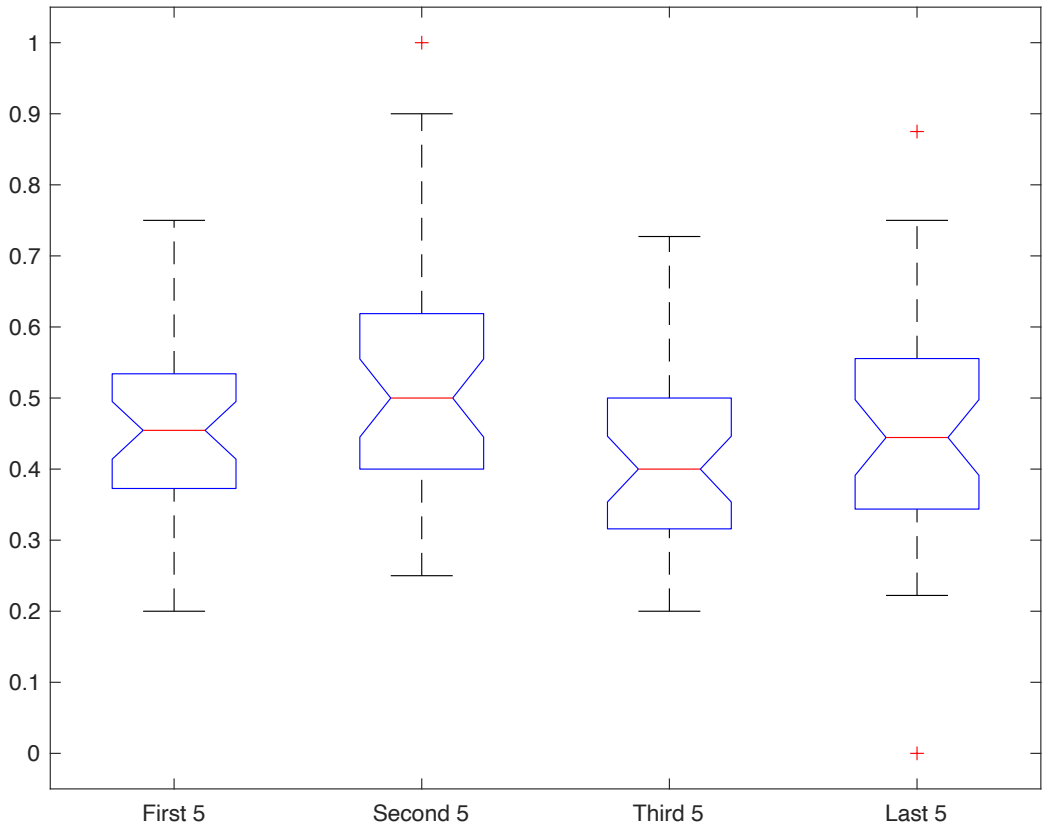


Figure 21. *Resultant box and whisker plot from an ANOVA analysis comparing the rate subjects chose the maximally information treatment assuming* ***Equation 12****. That is, whether subjects chose an option when its known AM equaled 0 and its unknown AM was greater than zero (resulting in an informativeness value of 1). Choice behavior did not significantly vary between 5-trial bins that an unknown DE was in-effect. Red markers indicate the median and box edges indicate the first (q_1) and third (q_3) quartiles of the sample of subject choice means in the analysis. Values beyond the whiskers (greater than $q_3 + 1.5 \times (q_3 - q_1)$ or smaller than $q_1 - 1.5 \times (q_3 - q_1)$ are outliers and indicate by red crosses.*

2.4. Discussion

This behavioral was conducted to investigate whether informativeness can be tracked using a quantifiable operational definition – at least more quantifiable than the categorical definitions found in the value-based decision neuroscience literature examining the explore-exploit dilemma. First, behavioral results suggested that subjects understood the task and learned to use weighted attributes to make high valued choices, as intended. That is, value-based decisions were dependent on the treatment values and not simply the attribute magnitudes (**Figure 18**). Further, these appetitive effects of the true treatment value significantly increased as the task progressed, an indication of learning (**Figure 17**). Specifically, choice accuracy was significantly lower and dependent more on the attribute magnitudes during the first five trials compared to the subsequent trials of the duration of an unknown drug effectiveness (**Figure 18**). It is proposed here that the dependence on attribute magnitudes during the first five trials was strategic and not random. The use of the attribute magnitudes could be described by a simple arithmetic definition of informativeness suggesting that subjects utilized the attributes during a phase where learning occurred. This phase occurred during the first five trials, as the subsequent trials found asymptotic performance, suggesting no learning occurred after the first five trials (**Figure 17**). Herein lies a limitation in the present thesis thus far: it is difficult to support that strategic decision making occurred (to facilitate learning) without gaze data or neuroimaging methods to estimating where attention was directed; such measures could contribute evidence for or against strategic decision making in the task (Fiedler et al., 2019; Gluth et al., 2020; Hunt et al., 2018). In any case, analyzing behavioral results alone suggested that subjects utilized attribute magnitudes in a manner described by **Equation 11** in a phase that coincided with learning in the task (**Figure 20**). One possible reason for this is that this manner removed the

known treatment value contributed by weighting the known attribute magnitude by the known drug effectiveness (note, terms as they related to the behavioral task are described in **Figure 8**) first, allowing subjects to more easily infer the unknown drug effectiveness from the left-over feedback (**Figure 15**). That is, it is speculated here that subjects performed the following computation to arrive at the unknown drug effectiveness value:

$$\frac{TV - known\ TV}{unknown\ AM} = unknown\ DE \qquad \text{[Equation 13]}$$

Subjects were more inclined to perform the above computation wherein a small numerator, as in the case when the known attribute magnitude is larger than the unknown attribute magnitude, is easier to learn about than a large numerator. An argument against using the known attribute magnitude this way were cases when it equaled zero. Therein, learning could not occur unless informativeness was based on the unknown attribute (**Equation 12**). In this case, when the known attribute equaled zero, informativeness would be instead be maximal. However, results showed that subjects did not significantly choose the option with the known attribute magnitude equal to zero during the first five trial bin, when information was preferred (**Figure 20**). This finding, and that information seeking was significant in the first bin, supports the description of exploratory behavior based on the known attribute magnitude (**Equation 12**).

In summary, subjects understood the behavioral task and performed as intended, deploying attribute-based, information-seeking decision making during learning phases – a sign of exploratory behavior. This preference for information and value changed a function of time, such that informativeness was influential during early-timepoints after the unknown drug effectiveness updated but then attenuated to insignificance on subsequent trials. At those subsequent phases after learning, value-based decision making was strictly deployed – a sign of exploitative behavior. These behavioral results contribute to the literature by demonstrating that

learning in a multi-attribute task demands for a deliberation between option value and informativeness, lending support to the attribute integration model of subjective value. Further and critically, informativeness of options can be quantifiable, where greater informativeness is more appetitive. However, the algorithmic underpinnings of the information-seeking decision making strategy has not been discussed. To be specific: what conditions, or latent variables, signal when to deploy exploratory and exploitative behaviors? The literature on these phenomena suggest it is the amount of uncertainty – this is maximal during periods when the drug effectiveness has update and must be re-learned. However, the next research questions investigates whether the quantifiable definition of informativeness can improve the ability of computational models to explain human decision making. In the next chapter, study two sought to examine this question by utilizing Bayesian modeling methods. These models can test the extent that information-seeking occurred for each subject and allowed examination this strategy was an optimizing or a hindering strategy of value-based decision making.

Chapter 3 – Bayesian Models

This chapter describes the computational modeling of simulated data and the behavioral findings in "**Chapter 2 – Behavioral Experiment**". Computational models are algorithmic hypotheses of behavior or cognition that allow researchers to probe non-linear and complex systems that are otherwise intractable with traditional or descriptive analyses (Farrell & Lewandowsky, 2018; Kriegeskorte & Douglas, 2018; Wilson & Collins, 2019). In an overview on the method, Wilson and Collins (2019) described a subset of applications of computational models. These applications are critical ideas to the approach in this chapter and so are briefly described in the first sub-chapter, in addition to their application in this chapter. In the following sub-chapter, a methods section describes the development of three Bayesian models that were used to explain human behavioral results in "**Chapter 2 – Behavioral Experiment**". Before briefly presenting these models, a critical variable needs to be discussed. The **expected value** (**EV**) of an option differs from that option's objective value (in the context of this thesis and behavioral experiment: the treatment value or TV; **Figure 8**). The TV is the objective true value of an option; the EV is the subjective value of an option, that integrates an estimate of relevant variables in the task (Glimcher, 2014). In this thesis, the two variables that influences the expected value are the estimate of the TV and the informativeness. These two terms should not be confused to be synonymous – they are distinct variables. Regarding the Bayesian models, each version is iteratively more complex, with the simplest model deliberating option EVs that only consider, a slightly more complex model caring about informativeness in addition to value, and a final complex model caring about informativeness and value as a function of uncertainty. Following the methods sub-chapter are model comparison results to identify the winning model and a discussion about the interpretation of these results and critical limitations.

3.1. Introduction

As mentioned, Wilson and Collins summarized the use and applications of computational in a 2019 review, describing four critical uses: simulations, parameter fitting, model comparison, and latent variable estimation. The first application is to simulate behavior; simulation methods analyze computer-generated data produced by a task-performing version of the computational model under study. The model parameters can be manipulated in order to simulate subjective differences between subjects – this enables mathematically grounded predictions of human performance in the real task. In addition to predictions, simulated data gives modelers a means to test whether their computational models and task paradigms are capable of measuring the psychological constructs under study (Farrell & Lewandowski, 2018; Wilson & Collins, 2019; for examples of simulation methods: Chau et al., 2014; Collins & Frank, 2012; Hunt et al., 2014; Rescorla & Wagner, 1972; Wang, 2002; Wang et al., 2018). To this end, simulation methods were applied in the present thesis to assess the behavior of the computational models under study. That is, using simulated data, the models returned the probability of choosing each treatment option in accordance with their EV (integrating their TV and informativeness). In doing so, human decisions could be understood subject-by-subject and mathematical based predictions can be made about human decision making. The second application of computational models is parameter fitting, or free parameter estimating, which allows researchers to account for differences of subjective variables that are theorized to underlie behavioral and cognitive constructs under study (Farrell & Lewandowsky, 2018; Wilson & Collins, 2019). There are a few methods to get these estimates, such as finding the maximum likelihood of the free parameter value given the data or by using optimization algorithms that test a myriad of rational values to see which describe, or fit, the data best (Cohen, 2017). The present thesis applied the latter

approach to estimate the degree of two free parameters: the first was related to how well subject decision making adhered to the true value of treatments; the second accounted for the extent of information-seeking expressed by each subject. More technically, models estimated a free parameter of value-based decision stochasticity and a second free parameter for the predilection towards information seeking (Farrell & Lewandowski, 2018; Wilson & Collins, 2019; for examples of parameter estimation discussed in this thesis: Chalk et al., 2010; Chau et al., 2014; Hunt et al., 2014; Levy & Glimcher, 2011; Ting et al., 2015). A third application of computational modeling is comparing different algorithmic hypotheses. This involves observing how well different models explain human or animal data relative to other models being tested. The best explanation of behavior is the winning model, and thus, the superior algorithmic hypothesis in the set of models under study (Farrell & Lewandowski, 2018; Wilson & Collins, 2019). In this thesis, the model comparison methods penalized for the number of free parameters in the model and considered the sample size – this benefitted simpler explanations of behavior (Akaike, 1974; Schwarz, 1978). In this thesis, the simpler explanation is the model without a free parameter for a predilection informativeness, which is the model that only cares about value, i.e., the null hypothesis, that informativeness is not considered during decision making, benefits in this thesis for being simpler. The fourth application is latent variable estimation, which allows researchers to estimate variables that evolve over time during behavior and cannot be captured by a static variable, such a free parameter (Bach & Dolan, 2012; Courville et al., 2006; O'Reilly, 2013; O'Reilly & Mars, 2015; for examples of latent variable estimation of uncertainty: Behrens et al., 2007; Courville et al., 2006; Meyniel & Dahaene, 2017; Muller et al., 2019; Nassar et al., 2010). Here, latent variable estimation was used to track subject estimates of the unknown DE (terms in **Figure 8** are used in this chapter as well) in the task and the amount of uncertainty

about those estimates on every trial. Bayesian inference was chosen to conduct these estimates because it is an optimal statistical procedure for imprecise decision making and trial-and-error learning (discussed in "**Chapter 1.3.2. – By Bayesian inference**"; O'Reilly, 2013; O'Reilly & Mars, 2015; Stone, 2013). A critical disclaimer regarding the use of Bayes' theorem in this thesis is warranted. As is common in the cognitive neuroscience literature, Bayesian inference is used here to compare and approximate inferential strategies implemented by humans in the behavioral experiment (Bowers & Davis, 2012b; Griffiths et al., 2012). In other words, it is used to approximate a subject's internal model of the task environment that is underlying their choices in the decision trials (**Figure 10**). The purpose was not to test whether humans are or are not Bayesian observers (discussed in the literature review). Instead, the statistical tool was used to approximate the role of trial-to-trial uncertainty in information-seeking versus value-based decision making.

3.2. Methods

Three versions of Bayesian models are described in this chapter. Generally speaking, these three differed in their preference for value and information and were composed of two central functions: a value function and an information function. The value function was based on Bayesian learning:

$$p(\theta \mid x_t) \propto p(x_t \mid \theta) * p(\theta) \qquad \text{[Equation 14]}$$

$$\text{Posterior} \propto \text{Likelihood} * \text{Prior}$$

In the context of this sub-chapter, θ is notation for a value that the unknown DE can take and x_t indicates an observation at trial *t*. A demonstration of how Bayesian inference was applied is shown in **Figure 22**. The plot displays the schedule of unknown DE during a simulated run of the task with 185 trials and the black- and green- dashed line indicates the estimate or inference

of the same unknown DE. Note that these simulation results only tested the Bayesian inference learner and so the update schedule and task duration are different than the actual behavioral experiment. **Figure 22** shows the learning patterns of the typical Bayesian learner and the evolution of uncertainty over time, with uncertainty indicated by the spread of the color map of each possible unknown DE over each trial – warmer colors indicating an increasing probability of the unknown DE being the true value. For instance, once the task began and when the unknown DE updated, the Bayesian observer became maximally uncertain. This is indicated by the high spread of warmer colors around the estimated (observed) unknown DE when the white line jumps to a new value in the statespace of possible unknown DEs. As more experience is gained, the Bayesian observer steadily reduces uncertainty after an updated unknown DE, indicated by a converging spread and warming of colors around the estimated value of the unknown DE; note the convergence towards the true value of the unknown DE (**Figure 22**). Predictions of human behavior can be extrapolated from these simulation results. For example, the unknown DE updates a little amount between trials 60 and 80. The model failed to detect an unknown DE update if the change in unknown DE is quantifiably subtle. Additionally, the model also predicts that with more experience, confidence in the estimated unknown DE reaches asymptote, regardless of the number of trials experienced (note estimates after trial 80 when the unknown DE no longer updates; **Figure 22**). Bayesian inference learning is special for this reason: it can mathematically infer a probable estimate of parameters while also keeping track of the amount of uncertainty in those estimates (**Figure 22**). This was utilized in the present thesis to examine human behavior in the behavioral experiment. Specifically, these components of Bayesian inference learning were critical to the computational models. The development, design, and results of these models are described in the upcoming sub-chapters.

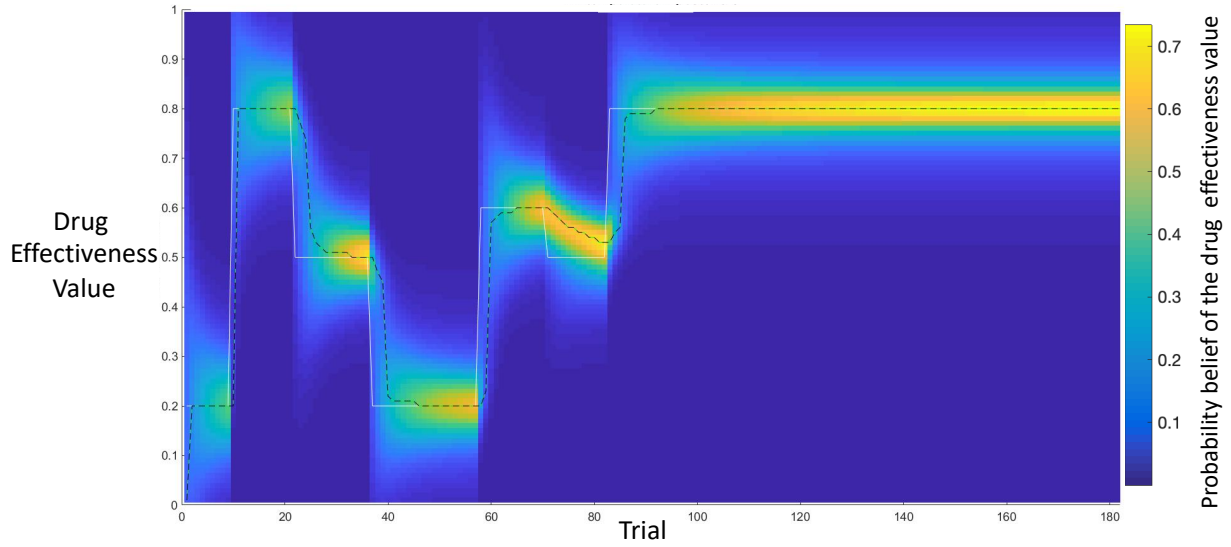


Figure 22. *A sample of the Bayesian observer learning the unknown DE in a similar task that human subjects performed. This figure plots the observer's beliefs, namely, the posterior probability distribution, over the course of the task. The horizontal axis indicates the trial; the vertical axis represents the statespace of the unknown DE, i.e. the possible values the unknown DE could take; cold colors indicate very low probability of the unknown DE at that trial of the corresponding value in the statespace; warmer colors indicate increasing probability that the corresponding value at that trial is the value of the unknown DE observed; color scale bar is shown on the right; the solid white line indicates the true value of the unknown DE on each trial; the dashed line indicates the running estimate of the unknown DE, i.e., the maximum a posteriori, at the corresponding trial.*

In summary, the computational models deploy Bayesian inference to estimate the unknown DE (**Equation 14**; repeated below for the convenience):

$$p(\theta \mid x_t) \propto p(x_t \mid \theta) * p(\theta)$$

The probability that the unknown DE equaled θ given an observation (the posterior probability) is proportional to the probability of a recent observation assuming that the unknown DE equaled θ (the likelihood) weighted by the probability of θ being the true value of the unknown DE (the prior probability); this latter computation is typically based on experience or some other heuristic. Finding the posterior for all possible values of the unknown DE gives a probability distribution that can be used to infer an estimate of the unknown DE (O'Reilly & Mars, 2015). The following discussion describes the implementation of these components of Bayesian learning and their integration to produce an estimate of the unknown DE after each trial.

3.2.1. Value

As mentioned above, the computational models had a value function and an information function. The value function is the learning algorithm of the models that estimates the unknown DE of a treatment by using feedback, AMs and the known DE in a decision trial, then uses this estimate to contribute to the expected value of future treatments (terms can be reviewed in **Figure 8**). The process to do this is described next. First, a vector whose elements represent all the possible values, i.e., the statespace, of the unknown DE was defined. This is notated by Θ and included real numbers between -1 and +1. Or notationally:

$$\Theta = (-1, 1) \qquad \text{[Equation 15]}$$

Note, the actual number of real numbers in this range is infinity, so a continuous account of this range must be approximated discretely to be analytically tractable. Approximating this range is easily implemented in MATLAB by creating a vector with the first element equal to -0.99, then iterating up to +0.99 by 0.01 as below (written as pseudocode for MATLAB):

$$statespace = [-0.99{:}\ 0.01{:}\ 0.99] \ = \ [-0.99, -0.98, -0.97, \dots, 0.97, 0.98, 0.99]$$

The resulting vector is the statespace of the unknown absorbance with 199 elements approximating a continuous distribution in **Equation 15**. At trial 1, subjects had no a priori information about the unknown DE. To represent this initial ignorance, the prior distribution was made uniform. These 'ignorant' distributions consider all the possible values in the statespace to be equally probable. In this case, the probability for each possible unknown DE, θ, in the statespace is as follows:

$$\mathrm{p}(\theta) = \frac{1}{elements\ in\ the\ statespace} = \frac{1}{199} \qquad \text{[Equation 16]}$$

A probability for each θ that makes up the statespace composes the prior probability distribution of the unknown DE; in the initial trial, this is a uniform distribution (for example, **Figure 3b,**

bottom right). The estimate of the unknown DE used by the value function is a product of the mean and standard deviation of this prior distribution. This is computed by taking the weighted sum of the values in the statespace with their corresponding probabilities in the prior distribution at trial 1 (O'Reilly, 2013). Notationally:

$$\theta_{\mathrm{est,t}} = \sum_{i=1}^{199} p_t(\theta_{\mathrm{i}}) * \theta_{\mathrm{i}} \qquad \text{[Equation 17]}$$

The estimated unknown DE at trial t, $\theta_{\mathrm{est,t}}$, is equal to the sum of possible values, approximated by the statespace with 199 elements, weighted by corresponding prior probabilities of each being the true DE. This estimation method is chosen over the alternate method of using the statespace element with the highest probability of being true, the maximum a posteriori (MAP). This latter method uses the peak of the distribution and is would optimize decision making if preference were based on this value. However, this estimate not account for the uncertainty, i.e., the standard deviation of the distribution, unlike the weighted sum (O'Reilly, 2013; O'Reilly & Mars, 2015). One potential flaw in using the weighted sum is that it starts with an estimate unknown DE of zero on trial 1 when the prior is a uniform distribution, introducing potential bias towards small absorbance values in the first trial. However, the effects of the prior become irrelevant over experience (Farrell & Lewandowsky, 2018; Stone, 2013): the uniform prior distribution results in the first posterior probability after trial 1 to be equal to the likelihood produced of the first observation. This likelihood function was Gaussian, with two parameters to define its distribution. The mean parameter was the observation after a choice and the standard deviation parameter was the standard deviation of the prior distribution. Notationally:

$$Likelihood\ distribution \sim N(x_i, std(\mathrm{p}(\Theta) * \Theta)) \qquad \text{[Equation 18]}$$

Thereafter, the likelihood distribution was normalized before weighting with the prior to produce the posterior distribution, per Bayes' theorem (**Equation 14**). In subsequent trials after the first, the prior distribution was equal to the posterior distribution produced by the previous observation. That is:

$$p(\Theta) \propto p(\Theta|\, x_{i-1}) \qquad \text{[Equation 19]}$$

The estimate of the unknown DE was obtained by computing the weighted mean of these prior distribution.

It is worth repeating that subjects did not receive warnings about updates to the unknown DE in the behavioral task. Further, subjects did not receive any hints about the schedule of these updates and were only notified that changes to the known and unknown DEs would occur at different separate times in the task. To capture the effects of anticipation of these updates in the model, a 'leak' was introduced to the estimated prior probability for each trial (Courville et al., 2006; Nassar et al., 2010; Wilson et al., 2010). This was formulated as follows:

$$p_{t+1}(\theta_i) = p_t(\theta_i|x_t) * (1 - H) + U(\theta) * H \qquad \text{[Equation 20]}$$

The notation *H* refers to the hazard rate, sometimes called the transition function, and indicates the probability that unknown DE, θ, had updated in the current trial (Courville et al., 2006; Eshel et al., 2015; Eshel et al., 2016; Muller et al., 2019; Nassar et al., 2010; O'Reilly, 2013; O'Reilly & Mars, 2015; Sarafyazd & Jazayeri, 2019; Wilson et al., 2010). Consequently, the above formulation states that the prior distribution for the upcoming trial *t+1* is equal to the running posterior distribution after a recent observation, x_t , and is weighted by the probability that a change has not occurred (*1 – H*). This is then summed to the uniform distribution weighted by the probability that a change has occurred (*H*, the hazard rate). In short, the above formulation is a concise computational representation of the belief that the variable that is being inferred, the

unknown DE, may be the same or different in the next trial by adjusting the running prior estimate towards an ignorant state, the uniform distribution. In the behavioral experiment, the unknown DE updated every 20 trials. This update was not signaled; therefore, it was expected that the true hazard rate needed to be inferred. To identify and formulate this learning, three assumptions were implemented as computational models beyond the three discussed in this chapter: in the first case, 1) subjects knew the true value of H, which is reasonable if subjects obtained enough information about the hazard rate from the known absorbance which changed at the same rate but offset to the updates of the unknown DE; alternatively, 2) each subject had a subjective estimate of the volatility of the task, in which case, this can be estimated by an optimization algorithm with the hazard rate as a free parameter; or 3) subjects learned the hazard rate by a similar mechanism as the unknown DE, i.e. a latent variable learned via Bayesian inference. Models implementing the second and third possibility did not significantly improve the ability of the value function to predict human choices. Therefore, the models discussed in this chapter adopted the simplest assumption of a flat hazard rate equal to the true probability of an update (i.e., equal to 1/20). This assumption and the above description of the value function were kept constant for the three models discussed in the present chapter.

3.2.2. Informativeness and uncertainty

The distinguishing feature of the three models compared here are their implementation of informativeness. The informativeness a treatment option was discussed in "**Chapter 2 – Behavioral Experiment**"; **Equation 11** repeated below for convenience:

$$informativeness\ (IF) = \frac{known\ AM}{known\ AM + unknown\ AM}$$

This estimate of informativeness was combined with the value function to approximate subjective preferences for the treatments on decision trials, i.e., they were combined to estimate

the deliberation of value and informativeness to produce the **expected value** (**EV**) of treatments. In the most complex model, this interplay between value and information was moderated by the amount of uncertainty about the running estimate of the unknown DE, generated by observations. The capability of Bayesian inference to estimate uncertainty in latent variables was deployed to this end, and could be formulated simply as below, described using pseudocode for MATLAB (a reminder that Θ represents the statespace defined in **Equation 15**):

$$uncertainty = std([\mathrm{p}(\Theta) * \Theta]) \qquad \text{[Equation 21]}$$

Model uncertainty was equal to the standard deviation of the probability distribution weighted by the corresponding statespace values – the spread of this product is the precision of the unknown DE, which is the definition of uncertainty. This ability of Bayesian inference is the reason it is attractive to researchers who want to investigate the effects of conscious imprecision during perceptual and preferential decision making (Bach & Dolan, 2012; Behrens et al., 2007; Cao et al., 2019; Chalk et al., 2010; Courville et al., 2006; Ernst & Banks, 2002; Frank et al., 2009; Gottlieb et al., 2020; Griffiths et al., 2012; Meyniel & Dahaene, 2017; Muller et al., 2019; O'Reilly, 2013; Nassar et al., 2010; Parr et al., 2018; Polanía et al., 2019; Radulescu et al., 2019; Stojić et al., 2020; Ting et al., 2015; Wilson et al., 2010).

3.2.3. Models and simulations

Ultimately, the goal is to identify the best algorithmic hypothesis of human behavior in the behavioral experiment by comparing three models deploying information-seeking differently. For brevity, the value function described above is abbreviated as VF. This was amended with an information function described above, abbreviated IF. Altogether, the general structural of the models estimates the EV for each option by combining the VF and IF as below:

$$EV = VF + k * IF \qquad \text{[Equation 22]}$$

The variable *k* is a free parameter that scales the *IF* (**Equation 11**) to make its value comparable with that of the *VF* – so that if this were true for a subject, they would then consider the possible reward of the option to be on par to its informativeness. So then, the free parameter *k* was necessary because *VF* had a maximum value of 16.2 – as in the case where both known and unknown DEs were positive 5-stars (see **Figure 9b** to see start-DE relationship) and their AMs were nine each (per **Equation 10**):

$$TV = (9 * 0.9) + (9 * 0.9) = 16.2$$

But *IF* has a maximum value of one – as in the case where the unknown AM equals zero (per **Equation 11)**:

$$\frac{known\ AM}{kown\ AM\ +\ 0} = 1$$

Therefore, the *k* variable in **Equation 22** can be described as an estimate of *information appetite*, indicating how important, relative to the *VF*, informativeness was to a subject during decision trials. As such, *k* was a free parameter: 0 was its lower bound and indicated that informativeness was irrelevant to a subject; 16.2 was the max value of the *VF* and was used as the upper bound for k, indicating that *IF* was as important as *VF* to subjects while they deliberated between options. The method of getting these estimates of *k* is described below in the sub-chapter on model fitting; it suffices to mention here that if informativeness was irrelevant to the subject, then the best fitting *k* value will be near zero, reducing the *EV* function in **Equation 22** to simply *VF*. But if subjects cared about informativeness as defined by **Equation 11**, then the *EV* function of **Equation 22** would best describe human performance in the task. In this latter case, the k free parameter will be greater than zero. Note, both formulations of *EV* described above (with *IF* and without) are static estimations of option values, and do not vary as a function of experience. For example, it is expected that subjects should be maximally uncertain at the start of the task or after

an update to the unknown DE and should attenuate with experience. The Bayesian learner captures this influence of uncertainty during the task. Specifically, the width of the probability distribution with respect to the unknown DE, i.e., its standard deviation, gives a measurement of precision in the estimated unknown DE.

This estimate of trial-by-trial uncertainty (**Equation 21**) about the unknown DE that subjects may have experience while learning was implemented in the EV function as follows:

$$EV = [(1 - uncertainty) * VF] + [uncertainty * \ k * IF] \quad \text{[Equation 23]}$$

Model uncertainty was computed by taking the square root of weighted variance of the statespace, weighted by the running prior distribution (**Equation 21**) – the same as the second parameter describing the likelihood distribution in the value function (**Equation 18**). After an update to the unknown DE was detected, model uncertainty spiked, and then decreased towards zero as the task proceeded (e.g., **Figure 22**). The EV function (**Equation 23**) predicts that an appetite for informativeness is maximal during high uncertainty, e.g., after the unknown DE update is detected resulting in more information-based decisions – exploratory behavior. This information appetite diminishes as model uncertainty approaches zero, resulting in more value-based decisions – exploitative behavior. The **Equation 23** model predicts that decisions would be influenced by both value and informativeness in a manner modulated by uncertainty: initially preference for information is greater than value and attenuates as the model becomes more certain in its estimate of the unknown DE.

The above describes the mathematical bases of the theories tested via modeling. The models are described next. The first model is the simplest, **baseBayes**. This is the case where EV is exclusively a function of estimated treatment value, therefore, this model is described as:

$$EV = VF \quad \text{[Equation 24]}$$

The next model, **infoBayes_v1**, considers informativeness of options but does not weigh the *VF* and *IF* with uncertainty. Therein, informativeness and value have a static preference throughout the task:

$$EV = [VF] + [k * IF] \quad \text{[Equation 25]}$$

The third model, **infoBayes_v2**, accounts for uncertainty, assuming value is greater during low uncertainty and informativeness is greatest during high uncertainty:

$$EV = [(1 - uncertainty) * [VF]] + [(uncertainty) * [k * IF]] \quad \text{[Equation 26]}$$

In order to test whether these models are in-line with predictions, simulations were conducted to confirm model choice patterns (Wilson & Collins, 2019). The task environment during simulations was similar to the environment subjects performed in, with the critical modification that one option was always the high-value, but low-information option and the alternative was the high-information but low-value option. If the models only sought value without considering information, the probability of the model choosing the high treatment value should always be at or above chance level (50% for two options) when the unknown DE update has been detected. Else, if the model seeks information when the unknown DE update is detected, the probability of choosing the high informative option should be greater than chance level. As mentioned, three models were tested. Model choice probabilities for the left and right option were estimated using the softmax function (**Equation 1**):

$$p_x(EV_x) = \frac{\exp\left(\frac{EV_x}{T}\right)}{\sum_{i=1}^{n} \exp\left(\frac{EV_i}{T}\right)}$$

Here again: *x* indicates any one option among an n number of options, EV_x indicates the expected value of *x*, and p_x indicates the probability of choosing option *x*. For the two alternate forced choice paradigms described in Chapter 2, the function above refers to the left option as '1' and

the right option as '2'. It states that the probability of choosing the left option is equal to the exponential of the expected value of the left treatment, EV_1, divided by the free parameter, *T*, also called the temperature, then this exponential is normalized for both displayed options in the trial to give the probability of a decision given the expected value. These are probabilities, so the sum of the probability for each option equals 1. Therefore, the probability of choosing the right option, EV_2, can be computed simply by 1 - EV_1 (and vice versa). As discussed in the chapter on value-based decision making, the *T* free parameter is an estimate of the stochasticity of subject behavior, where low *T* values near 0 indicate value-based, deterministic decisions and higher T values indicate more random behavior (Farrell & Lewandowsky, 2018; Wilson & Collins, 2019). This parameter is commonly used in the decision neuroscience literature to estimate the accuracy of subjects' assessments, i.e., their predilection for choosing the high valued options (for examples: Ballard et al., 2018; Chau et al., 2015; Collins & Frank, 2012; Hunt et al., 2012; Jocham et al., 2012; Leong et al., 2017; Niv et al., 2015; Shiner et al., 2012). The decision patterns that these models predict can then be compared with human behavior. The MATLAB scripts implemented for simulations were adapted to analyze and fit models of human behavior, as will be described below. For clarity, the models described in this main body are summarized in **Table 1** below.

Table 1. *The computational models under study.*

Model Name	Value Function (VF)	Information Function (IF)	Expected Value (EV)
baseBayes	p(Θ|x) weighted sum		[VF]
infoBayes_v1	p(Θ|x) weighted sum	kn_attr / (unkn_attr + kn_attr)	[VF] + [**k** * IF]
infoBayes_v2	p(Θ|x) weighted sum	kn_attr / (unkn_attr + kn_attr)	[(1 - σ) * VF] + [σ * (**k** * IF)]

3.2.4. Model fitting and comparison

Human performance in the behavioral experiment was compared with the choice preferences of the three models in **Table 1**. The goal was to identify the extent that these models,

algorithmic hypotheses of value-based and information-seeking in the task, are able to predict decision making and compare their explanatory power in order to, ultimately, identify the winning model – the model that best describes human decision making. To restate the aforementioned models, if humans do not use information during the task and just care about their estimates of the treatment values (TVs; terms are described in **Figure 8**), then the **baseBayes** model should perform best; else, if information is a factor during decision making and statically competes with value, then the **info_Bayes_v1** should perform better than the **baseBayes** model; further still, if subjects used information to the extent they were uncertain, such as when the unknown DE recently updated, then the **infoBayes_v2** model should outperform the **info_Bayes_v1** and **baseBayes** models. This sub-chapter describes how model fitting and comparison were used to identify the best fitting model of human performance in the task described in "**Chapter 2 – Behavioral Experiment**".

First, choice probabilities were obtained using a softmax function in the same manner as in the simulations. After choice probabilities were generated for all the options in each trial, the probabilities of the actually chosen options were used to compute log likelihood, a measure of how well the model described behavior. The measurement depends on parameter values and model under study (Farrell & Lewandowsky, 2018; Wilson & Collins, 2019). Notationally:

$$LL = \sum log\ p(EV_{chosen} | \ parameters, model) \qquad \text{[Equation 27]}$$

In short, the log likelihood is the probability of observing the data given the model and its parameters. In this case, the data are the choices made by subjects and the model was either **baseBayes**, **infoBayes_v1**, or the **infoBayes_v2**. The number of free parameters varied between the **baseBayes** and the information models (either **infoBayes_v1** or **infoBayes_v2**; **Table 2**). All models had a temperature, *T*, for the softmax function (**Equation 1**). The free parameter was free

to take any real number between 0 and 100. The information models included a second free parameter, *k,* which scaled the information function (discussed in "**3.2.3. Models and simulations**") to make informativeness comparable with value. This free parameter could take any real number between 0 and 16.2; the upper bound was chosen because it was the max value of the value function.

Table 2. The free parameters of the models under study. *Treatment values are learned and estimated by taking the means of probability distributions of the unknown DE generated by Bayesian inference. This property is shared across models. Informativeness is considered by two of the models the three models, infoBayes_v1 and infoBayes_v2. Additionally, uncertainty (**Equation 21**) factors into the infoBayes_v2 model.*

	Value Function	Info Function	Free Parameters
baseBayes	Bayesian latent variable		T
infoBayes_v1	Bayesian latent variable	known attribute/(known + unknown attributes)	T, k
infoBayes_v2	(1-σ) * Bayesian latent variable	σ * known attribute/(known + unknown attributes)	T, k

The free parameter values that produced the smallest log likelihood served as indicator of fit. The free parameters were obtained using MATLAB's optimization algorithm, *fminsearch*, to seek global minima solutions of the models under study (Cohen, 2017). Notably, the log likelihood is computed using a sum of probabilities of the selected options. As these are probabilities, with values less than 1 entered into **Equation 27**, the sum is a negative value. Note the following limitation: the best fitting model will have the smallest negative magnitude, but *fminsearch* searches for the minimum which would lead it to choose values towards negative infinite (Cohen, 2017). A blunt solution, used here, is to feed the optimization algorithm the negative of the log likelihood. A second limitation of optimization algorithms is that they identify local minima, with no guarantee that these minima are the desired global minimum. To overcome this limitation, *fminsearch* function was conducted with several starting values within the bounded range of values of the respective free parameters. These methods were adopted from published applications in the fields of perceptual and preferential decision neuroscience (Wilson & Collins, 2019; for examples: Chalk et al., 2010; Chau et al., 2014).

Model comparison was performed using two criteria: the **Akaike Information Criterion**, herein the **AIC** (Akaike, 1974) and the **Bayes Information Criterion**, herein the **BIC** (Schwarz, 1978), each respectively formulated below:

$$AIC = -2\,x\log(LL) + 2\,x\,numParameters \quad \text{[Equation 28]}$$

$$BIC = -2\,x\log(LL) + numParameters\,x\log(numObservations) \quad \text{[Equation 29]}$$

The AIC accounts for the log likelihood and the number of free parameters, whereas the BIC further accounts for the number of observations, making the model comparison more sensitive to the number of parameters. Each criterion assigns each model a score. The model with the smallest score, i.e., approaching negative infinity, is the winning model (Akaike, 1974; Farrell & Lewandowsky, 2018; Schwarz, 1978; Wilson & Collins, 2019). The BIC penalizes models for the number of parameters so as to further avoid a type II error, but otherwise this and the AIC are both commonly used in the cognitive neuroscience literature for model comparison (Farrell & Lewandowsky, 2018; Wilson & Collins, 2019). To be comparable with the literature, both scores are reported below.

3.3. Results

First, the simulation results are presented first. Simulations demonstrated the preference patterns predicted by the computational models while deliberating between two options during simulated decision trials. The first option was high in value but low in informativeness; the alternate option was low in value but high in informativeness. These simulated decision trials are intentionally different than the decision trials in the behavioral task, wherein option value and informativeness varied pseudo-randomly (**Figure 16**). This allowed plotting simulation results concisely in **Figure 23**: if subjects prefer value over informativeness, the probability “P(high chosen)” will be

greater than chance level of 50%. Conversely, if subjects prefer informativeness over value, then "P(high chosen)" will fall below 50%. The simulation results in **Figure 23** are described below.

baseBayes

The computerized chooser quickly learned to choose the high treatment value (**Figure 23a)**. If subjects were Bayesian optimal in the task, their performance would look like the **baseBayes** results in **Figure 23a**. Therein, after the first trial, the model developed a near-perfect estimate of the unknown DE. Notably, when the update was detected, the model was uncertain but did not seek the informative choice, as evident by the chance level probability of choosing the high treatment value soon after the DE updated. After only one trial from task onset or unknown DE update, the **baseBayes** model chose the high treatment option with over 90% accuracy. Additionally, after a few more trials, accuracy plateaued near 99% accuracy (**Figure 23a**). Whether subject performance adhered to this near-perfect accuracy depends on the temperature parameter of the softmax function, wherein high temperature may cause subjects to rarely choose the high treatment option. In **Figure 23**, the temperature was held constant

infoBayes_v1

This model utilized informativeness (**Figure 23b**). After task onset or an unknown DE update, the probability of choosing the high treatment value dropped well below 50% for the first trial, indicating a strong preference for the high informative treatment. Notably, this consideration of informativeness appears to distract from the accuracy value-based decision

making, indicating by diminished choice accuracy overall and a lower plateaued accuracy value to 91% given the same temperature free parameter.

infoBayes_v2

This model weighed informativeness and value by uncertainty (**Figure 23c**), evidenced by a diminished slope during the first five trials after task onset or unknown DE update. This indicated that the model initially preferred informativeness during a period when it was highly uncertain, as when the task begins, or the unknown DE has updated. Notably, choice accuracy is further diminished, plateauing near 80% accuracy given the same free parameter values as the previous model, **infoBayes_v1**. This is due to the inclusion of the uncertainty as a latent variable in the model (**Equation 26**).

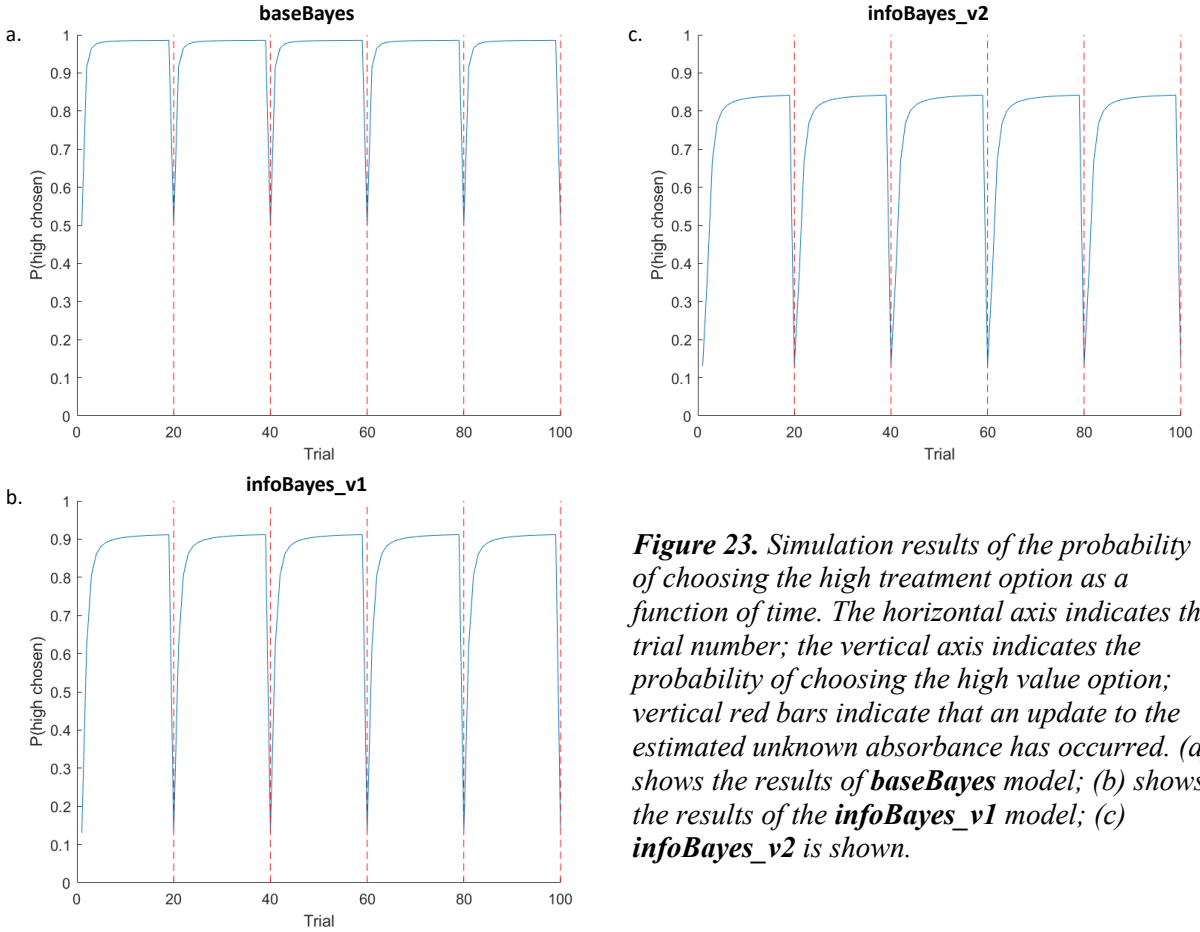


Figure 23. *Simulation results of the probability of choosing the high treatment option as a function of time. The horizontal axis indicates the trial number; the vertical axis indicates the probability of choosing the high value option; vertical red bars indicate that an update to the estimated unknown absorbance has occurred. (a) shows the results of* ***baseBayes*** *model; (b) shows the results of the* ***infoBayes_v1*** *model; (c)* ***infoBayes_v2*** *is shown.*

Model Fitting and Comparison

Next, model fitting and comparison results were conducted with an optimization algorithm by MATLAB to obtain the best fitting parameters of the models. This optimization function was conducted for each by subject (Cohen, 2017). Then, MATLAB was used to compute the AIC (**Equation 28**) and BIC (**Equation 29**) scores for model comparison. The **infoBayes_v2** model was the winning model (**Figure 24**) given its significantly lower AIC score across subjects (mean = 153.2 $\pm$ 7.54; **Figure 24, left**) and BIC scores (159.8 $\pm$ 7.54; **Figure 24, right**) relative to results of the **baseBayes** model (AIC mean = 161.3 $\pm$ 7.48; BIC = 164.6 $\pm$ 7.48; conducting a t-test comparing **baseBayes** with **infoBayes_v2** yields the following comparisons: $t_{AIC}(38) = 3.75$, $p_{AIC} = 0.0006$; $t_{BIC}(38) = 2.24$, $p_{BIC} = 0.03$) and comparison with **inforBayes_v2** with the **infoBayes_v1** model (AIC mean = 158.0 $\pm$ 7.32; BIC = 164.6 $\pm$ 7.32) yields the following t-test results: $t_{AIC}(38) = 3.94$, $p_{AIC} = 0.0003$; $t_{BIC}(38) = 3.94$, $p_{BIC} = 0.0003$). To summarize the above (a reminder that lower scores indicate better fit):

AIC results: infoBayes_v2 < infoBayes_v1 < baseBayes

BIC results: infoBayes_v2 < infoBayes_v1 = baseBayes

The results comparing the **baseBayes** with the **infoBayes_v1** models were less consistent. AIC results suggested that the **infoBayes_v1** was the runner-up model but the BIC comparison indicated no difference between these model fits (t-tests comparing **baseBayes** with **infoBayes_v1**: $t_{AIC}(38) = 2.31$, p_{AIC} 0.03; $t_{BIC}(38) = 0.047$, $p_{BIC} = 0.96$). A reminder that an information appetite parameter *k* was computed for each subject with MATLAB. This allowed a straightforward and final comparison of information-seeking with performance of human subjects in the behavioral experiment. A moderate and significant correlation was observed between the value of the information appetite and the rate of choosing the high value option in

the task ($r = -0.63$, $p < 0.01$): greater accuracy results in the task significantly correlated with a lower information appetite.

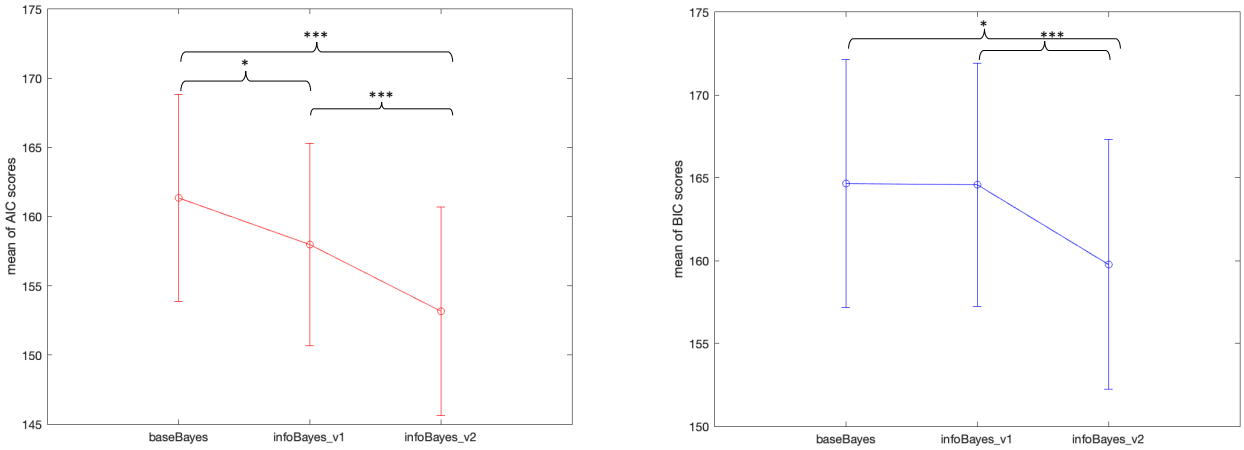


Figure 24. *Model comparison results. One asterisk indicates significant difference with $p < 0.05$ but $p > 0.01$; three asterisks indicate $p < 0.01$. Lower AIC and BIC scores indicate better model fit.*

3.4. Discussion

Bayesian modeling was used to track learning and preference for option informativeness and value and in the multi-attribute bandit task described in "**Chapter 2 – Behavioral Experiment**". Additionally, informativeness was operationally defined as a quantitative variable, a contrast from the decision neuroscience literature wherein an option is or is not and informative option (sub-chapter "**1.5. Informativeness**"). The first model, **baseBayes**, had no preference for informativeness, instead basing decisions only on value. The second model, **infoBayes_v1**, accounted for informativeness but assumed its influence was static, only based on the attributes. That is, this model had a constant preference for value and informativeness throughout the task, not adapting to how much it has learned and its own imprecision. The third model, **infoBayes_v2**, preferred value and informativeness as a function of uncertainty. That is, when uncertainty was highest after the unknown DE updated, the model preferred informative options, but this appetite for information attenuated with experience. The **infoBayes_v2** was the winning model, being the best descriptor of human decision making in the task. As observed in the

behavioral analysis, subjects utilized the attribute magnitudes during the first five trials that an unknown DE was in effect in a manner unrelated to treatment values. The **infoBayes_v2** model suggests this manner was information seeking – when subjects were the most uncertain, this winning model significantly preferred informative options, then it reverted to value-based decision making exclusively following more experience. In the literature the former strategy is termed exploratory behavior and the latter termed exploitative behavior – the results suggest that both of these strategies are driven by quantitative processes. Exploration assessed informativeness in a quantitative manner similar to the assessment of value, as opposed to a categorical assessment of information implied by the literature on information seeking. Notably, in this thesis, information seeking predicted low performance. It is speculated here that an appetite for information was higher when subjects found the task more difficult, prolonging the need for information-based decision making compared to subjects who found the task easy. Indeed, mathematically savvy subjects would be able to learn the value of the unknown DE after a single trial when the unknown attribute did not equal zero, leading to a smaller information appetite. To get around this potential confound, future versions of the task may base stimuli attribute magnitudes on visual properties, such as volume or color intensity, instead of numbers. But it should be noted that if this confound was significant across subjects, then the unknown DE would be immediately detectable. In which case, the probability distributions estimating the unknown DE would be composed of zeros except for a singularity over the observed outcome (Farrell & Lewandowsky, 2018). However, neither human performance on decision trials nor survey trials approached this deterministic learning. Further, the **baseBayes** model would have been the winning model because it only considers the running estimate of the treatment value and ignores informativeness during option deliberation. Therefore, although deterministic

learning was possible, the results suggest learning was trial-and-error based, wherein information- and value-based decisions were deliberated during the deployment of exploratory or exploitative decision making. Ultimately, these results contribute to the literature by using Bayesian models in a multi-attribute value-learning environment and the attribute integration model of subjective value to enlighten the quantitative nature of exploratory behavior.

4. General Discussion

The present thesis utilized theories and methods from the cognitive neuroscience literature to simulate and test algorithmic hypothesis of information- and value-based human decision making. Specifically, the attribute integration model of subjective value (Rangel & Clithero, 2014) was used to develop a novel, computerized behavioral task for human subjects to examine the influence of option informativeness, operationally defined as a quantitative variable and manipulated alongside option value, on preferential decision making. Indeed, this quantitative operation definition contrasts with the typical categorical definition observed in the relevant studies of information seeking during the explore-exploit dilemma (Frank et al., 2009; Trudel et al., 2020; Wilson et al., 2014). That is, it was observed that increasing option informativeness results in increasingly more appetitive options to subjects, subsequently increasing the probability that the more informative option would be selected, independent of the value (**Figure 20**). This was done using traditional descriptive statistics and Bayesian modeling, the latter of which allowed further examination of for a potential explanation about the periods when information was significantly appetitive. Subjects preferred informativeness during periods of high uncertainty, then updating their preferences to value exclusively after the learning period was finished (a pattern captured by **Equation 26**). In these analyses, informativeness was defined by an arithmetic expression: the known attribute magnitude of an option divided by the sum of all its component attributes magnitudes. This definition of informativeness in the task was straightforward given the use of numerical stimuli determining option values. Another critical observation is that subjects initially pursued more informative options after the unknown drug effectiveness updated but not after more experience with this same drug effectiveness. This was not the case when the known drug effectiveness updated, suggesting subjects understood the

task and were responding to increased uncertainty with exploratory behavior. Notably, the use of informativeness as described above was akin to noise reduction. That is, during the period where subjects preferred informativeness, the options with a higher known attribute magnitude were significantly preferred (**Figure 20**). This had the effect of minimizing the contribution of the unknown drug effectiveness to feedback during a period when its estimated drug effectiveness was the least precise. Given the nature of the task, it is speculated here that this made the task easier for the subjects: subjects could infer the unknown drug effectiveness with better precision when its weighted value was smaller (**Figure 20**). Traditional statistics (**Figure 17**) and Bayesian modeling results supported this interpretation that the present definition of informativeness was in-fact related to the effects of information discussed in the sub-chapter "**1.5. Informativeness**". Information as defined was sought during uncertain periods. The best fitting model for behavior deliberated value and informativeness during preferential choice, further considering informativeness when uncertainty was high, like during early periods after an unknown drug effectiveness updated. Following experience, value-based decision making was the preferred strategy over information-based decision making.

However, there are several limitations and potentially problematic model design choices in this thesis. First, the estimate of the unknown drug effectiveness generated by the Bayesian models were not directly affected by informativeness itself, as defined in the literature review (**Figure 4**). It is expected that option informativeness should accelerate learning – however, these influences were not implemented by the model. A second limitation is the indirectness of the operational definition of informativeness used in this thesis (**Equation 11**). As defined, information is the quality that reduces uncertainty and is most simply represented by **Equation 12**. However, the behavioral results indicated that this strategy was not deployed by subjects in

the task (**Figure 20** and **Figure 21**), with patterns of information seeking better captured by **Equation 11**. A third concern is a mathematical ability confound. Subjects who are mathematically savvy may find the task easy relative to non-math savvy subjects. However, as described in "**Chapter 1 – Bayesian Models**", such an occurrence would have manifested in perfect performance after the first five trials, which was not case for any subject. Further, "**Chapter 2 – Bayesian Models**" would have identified this perfect performance as a singularity occurrence, which was not the case for any subject. Regardless, this problematic confound could be resolved by converting the numerical presentation of the stimuli into a visual representation, such as varying the volume or shape size of stimuli to indicate the same schedule of attribute magnitudes and drug effectiveness. A fourth limitation is the focus on behavioral data and lack of supplementary neural activity measure, such as fMRI, or attention estimate, such as eye tracking. However, as it has been argued by Niv (Forthcoming), understanding the computational problems of behavior and cognition or quintessential to understanding how the human brain works. The focus on developing a paradigm and computational model in this paper are significant in this regard; additionally, the behavioral task was developed in preparation for these traditionally neuroscientific research endeavors – as they are easy to adapt to such investigations. A fifth limitation in this thesis regards a common issue with research comparing different computational models (Griffiths et al., 2012; Wilson & Collings, 2019). It was possible that a superior information-seeking or value-seeking-only model not considered here exists or can be designed. For example, models based on Polanía's and colleagues' (2019) proposed method of computing choice predictions according to Bayesian inference (as opposed to the softmax function; **Equation 1**) or extending sequential sampling theories (Busemeyer et al., 2019). This issue was not acknowledged during the analysis of the three models presented; the goal here was

simply to investigate the influence of common estimate of uncertainty generated via Bayesian inference. Future investigations should observe whether alternate measures do better to test these hypotheses. A sixth issue with this thesis is the justified concern that the literature may not need a new cognitive task to investigate the effects of uncertainty on information seeking. For instance, Wilson and colleagues (2014) developed an oft used and clever paradigm they called the 'horizon task' to investigate directed exploration and the effect that varying the number of decision opportunities has on the willingness to explore; Frank and colleagues (2009) used a paradigm called the 'temporal utility integration task' to investigate information-based decisions that confirmed the appetitive or aversive quality of options; the famous and clinically critical 'Wisconsin Card Sorting task' can yield data that can be analyzed in the context of information-seeking, allowing for clear generalizability. As discussed in "**1.5. Informativeness**", these paradigms in the literature operationally define an information-seeking choice categorically: a choice in the task is either value-based or information-based. In the present thesis, informativeness was defined as a more quantitative variable in a similar range as value. This operational definition is a novelty in the research on the neuroeconomics of the explore-exploit dilemma, with the aforementioned paradigms being unable to answer the research questions presented in sub-chapter "**1.6. Research gap and questions**". Therefore, an important contribution to the literature is the behavioral task itself as the potential to explore the following question: does informativeness have the same qualities as reward amount? That is, much has been shown about the effects of reward amount on preference, such as ceiling effects on appetitive qualities and the over-sensitivity to aversive outcomes (Glimcher, 2014). Do these same properties describe informativeness, the quality that reduces uncertainty? This is a critical

line of investigation that is possible to explore with paradigms like the one developed for the present thesis.

In summary, the illumination of the explore-exploit dilemma is critical to our understanding of how the human brain works. Indeed, it has been suggested that this dilemma, as a computational problem (Kriegeskorte & Douglas, 2018; Marr, 1982), has guided the evolution of the granular prefrontal cortex (Passingham & Wise, 2012). This computational problem was investigated using a novel multi-attribute bandit task and Bayesian model analysis in the present thesis. First, the behavioral task allowed investigation for whether informativeness can defined as a quantifiable variable, as opposed to the literature that defines it as a categorical operational definition. Indeed, this quantifiable variable was able to capture a typical pattern of exploratory behavior found in the literature. Second, Bayesian modeling allowing the investigation of a potential hypothesis for the reasons underlying the patterns of exploration – namely, the influence of uncertainty in the deliberation of value and informativeness. There are further questions about informativeness to explore, but this thesis presents a means of investigating and exploring this critical construct mathematically.

References

Akaike H (1974). A new look at the statistical model identification. IEEE Transactions on Automatic Control, 19, 716-723.

Bach DR & Dolan RJ (2012). Knowing how much you don't know: a neural organization of uncertainty estimates. Nature Reviews, 13, 572-586.

Bartra O, McGuire JT, & Kable JW (2013). The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. Neuroimage, 76, 412-427.

Ballard I, Miller EM, Piantadosi ST, Goodman ND, & McClure SM (2017). Beyond Reward Prediction Errors: Human Striatum Updates Rule Values During Learning. Cerebral Cortex, 28, 3965-3975.

Beck JM, Ma WJ, Kiani R, Hanks T, Churchland AK, Roitman J, Shadlen MN, Latham PE, & Pouget A (2008). Probabilistic Population Codes for Bayesian Decision Making. Neuron, 60, 1142-1152.

Beesley T, Nguyen KP, Pearson D, & Le Pelley (2015). Uncertainty and predictiveness determine attention to cues during human associative learning. The Quarterly Journal of Experimental Psychology, 68, 2175-2199.

Behrens TEJ, Woolrich MW, Walton ME, & Rushworth MFS (2007). Learning the value of information in an uncertain world. Nature Neuroscience, 10, 1214-1221.

Bilder RM, Volavka J, Lachman HM, & Grace AA (2004). The catechol-O-methyltransferase polymorphism: Relations to the tonic-phasic dopamine hypothesis and neuropsychiatric phenotypes. Neuropsychopharmacology, 29, 1943-1961.

Boccara CN, Nardin M, Stella F, & Csicsvari J (2019). The entorhinal cognitive map is attracted to goals. Science, 363, 1443-1447.

Bowers JS & Davis CJ (2012a). Bayesian Just-So Stories in Psychology and Neuroscience. Psychological Bulletin, 138, 389-414.

Bowers JS & Davis CJ (2012b). Is That What Bayesians Believe? Reply to Griffiths, Chater, Norris, and Pouget. Psychological Bulletin, 138, 423-426.

Brunel N & Wang XJ (2001). Effects of Neuromodulation in a Cortical Network Model of Object Working Memory Dominated by Recurrent Inhibition. Journal of Computational Neuroscience, 11, 63-85.

Bryant RA & Nickerson A (2013). Treatment of Complex PTSD: The Case of a Torture Survivor. In O'Donohue W & Lilienfeld SO (editors). Case studies in clinical psychological science: Bridging the gap from science to practice. Oxford University Press.

Burke CJ, Soutschek A, Weber S, Beharelle AR, Fehr E, Haker H, & Tobler PN (2018). Dopamine Receptor-Specific Contributions to the Computation of Value. Neuropsychopharmacology, 43, 1415-1424.

Busemeyer JR, Gluth S, Rieskamp J, & Turner BM (2019). Cognitive and Neural Bases of Multi-Attribute, Multi-Alternative, Value-Based Decisions. Trends in Cognitive Sciences, 23, 251-263.

Calabresi P, Picconi B, Tozzi A, & Filippo MD (2007). Dopamine-mediated regulation of corticostriatal synaptic plasticity. Trends in Neuroscience, 30, 211-219.

Cao Y, Summerfield C, Park H, Giordano BL, & Kayser C (2019). Causal Inference in the Multisensory Brain. Neuron, 2012, 1076-1087.

Chalk M, Seitz AR, & Seriès P (2010). Rapidly learned stimulus expectations alter perception of motion. Journal of Vision, 10, 1-18.

Chau BKH, Jarvis H, Law CK, & Chong TT (2018). Dopamine and reward: a view from the prefrontal cortex. Behavioural Pharmacology, 29, 569-583.

Chau BKH, Kolling N, Hunt LT, Walton ME, Rushworth MF (2014). A neural mechanism underlying failure of optimal choice with multiple alternatives. Nature Neuroscience, 17, 463-70.

Chau BKH, Law CK, Lopez-Persem A, Klein-Flügge MC, & Rushworth MFS (2020). Consistent patterns of distractor effects during decision making, 9, 1-36.

Chau BKH, Sallet J, Papageorgiou GK, Noonan MAP, Bell AH, Walton ME, & Rushworth MFS (2015). Contrasting Roles for Orbitofrontal Cortex and Amygdala in Credit Assignment and Learning in Macaques. Neuron, 87, 1106-1118.

Chiba T, Kanazawa T, Koizumi A, Ide K, Taschereau-Dumouchel V, Boku S, Hishimoto A, Shirakawa M, Sora I, Lau H, Yoneda H, Kawato M (2019). Current status of neurofeedback for post-traumatic stress disorder: A systematic review and the possibility of decoded neurofeedback. Frontiers in Human Neuroscience, 13, 1-13.

Cohen MA, Dennett DC, & Kanwisher N (2016). What is the Bandwidth of Perceptual Experience? Trends in Cognitive Sciences, 20, 324-335.

Cohen, MX (2017). MATLAB for brain and cognitive scientists. Cambridge, Massachusetts: The MIT Press.

Collins AGE & Frank MJ (2012). How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. European Journal of Neuroscience, 35, 1024-1035.

Collins AGE & Frank MJ (2013). Cognitive Control Over Learning: Creating, Clustering, and Generalizing Task-Set Structure. Psychological Review, 120, 190-229.

Cools R & D'Esposito M (2011). Inverted-U-Shaped Dopamine actions on Human Working Memory and Cognitive Control. Biological Psychiatry, 69, 113-125.

Cools R, Gibbs SE, Miyakawa A, Jagust W, & D'Esposito M (2008). Working Memory Capacity Predicts Dopamine Synthesis Capacity in the Human Striatum. The Journal of Neuroscience, 28, 1208-1212.

Cools R, Frank MJ, Gibbs SE, Miyakawa A, Jagust W, & D'Esposito M (2009). Striatal Dopamine Predicts Outcome-Specific Reversal Learning and Its Sensitivity to Dopaminergic Drug Administration. The Journal of Neuroscience, 29, 1538-1543.

Cools R & Robbins TW (2004). Chemistry of the adaptive mind. Philosophical Transactions: Series A, Mathematical, physical, and engineering sciences, 362, 2871-2888.

Constantinescu AO, O'Reilly JX, Behrens TEJ (2016). Organizing conceptual knowledge in humans with a gridlike code. Science, 352,1464–1468.

Courville AC, Daw ND, & Touretzky DS (2006). Bayesian theories of conditioning in a changing world. Trends in Cognitive Sciences, 10, 294-300.

Crockett MJ & Fehr E (2014). Pharmacology of economic and social decision-making. In Glimcher PW & Fehr E (Editors), Neuroeconomics: Decision making and the brain (2nd edition). London, United Kingdom: Elsevier Academic Press.

D'Esposito M & Postle BR (2015). The Cognitive Neuroscience of Working Memory. Annual Review of Neuroscience, 66, 115-142.

Daw ND & Tobler PN (2014). Value Learning through Reinforcement: The Basics of Dopamine and Reinforcement Learning. In Glimcher PW & Fehr E (Editors), Neuroeconomics:

Decision making and the brain (2nd edition). London, United Kingdom: Elsevier Academic Press.

de Leeuw, JR (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. Behavior Research Methods, 47, 1-12.

Dolan RJ & Dayan P (2013). Goals and Habits in the Brain. Neuron, 80, 312-325.

Doll BB, Bath KG, Daw, ND, & Frank MJ (2016). Variability in Dopamine Genes Dissociates Model-Based and Model-Free Reinforcement Learning. The Journal of Neuroscience, 36, 1211-1222.

Doll BB, Hutchison KE, & Frank MJ (2011). Dopaminergic Genes Predict Individual Differences in Susceptibility to Confirmation Bias. The Journal of Neuroscience, 31, 6188-6198.

Dorris MC & Glimcher PW (2004). Activity in posterior parietal cortex is correlated with the relative subjective desirability of action. Neuron, 44, 365-378.

Dutilh G & Rieskamp J (2016). Comparing perceptual and preferential decision making. Psychonomic Bulletin and Review, 23, 723–737.

Ebitz RB, Albarran E, & Moore T (2018). Exploration Disrupts Choice-Predictive Signals and Alters Dynamics in Prefrontal Cortex. Neuron, 97, 450-461.

Echeveste R, Aitchison L, Hennequin G, & Lengyel M (2020). Cortical-like dynamics in recurrent circuits optimized for sampling-based probabilistic inference. Nature Neuroscience.

Edelman S (2008). Computing the mind: how the mind really works. Oxford University Press, New York.

Elton A, Smith CT, Parrish MH, & Boettiger CA (2017). COMT Val158Met Polymorphism Exerts Sex-Dependent Effects on fMRI Measures of Brain Function. Frontiers in Human Neuroscience, 11, 1-11.

Ernst MO & Banks MS (2002). Humans integrate visual and haptic information in a statistically optimal fashion. Nature, 415, 429-433.

Eshel N, Bukwich M, Rao V, Hemmelder V, Tian J, & Uchida N (2015). Arithmetic and local circuitry underlying dopamine prediction errors. Nature, 525, 243-246.

Eshel N, Tian J, Bukwich M, & Uchida M (2016). Dopamine neurons share common response function for reward prediction error. Nature Neuroscience, 19, 479-486.

Farrell S & Lewandowsky S (2018). Computational Modeling of Cognition and Behavior. Cambridge: Cambridge University Press.

Fellows, LK (2006). Deciding how to decide: ventromedial frontal lobe damage affects information acquisition in multi-attribute decision making. Brain, 129, 944-952.

Fiedler S, Ettinger U, & Weber B (2019). Neuroeconomics. In Klein C Ettinger U (Editors), Eye Movement Research. Studies in Neuroscience, Psychology and Behavioral Economics. Springer, Cham.

Filla I, Bailey MR, Schipani E, Winiger V, Mezias C, Balsam PD, & Simpson EH (2018). Striatal dopamine D2 receptors regulate effort but not value-based decision making and alter the dopaminergic encoding of cost. Neuropsychopharmacology, 43, 2180-2189.

Fiorillo CD, Tobler PN, & Schultz W (2003). Discrete Coding of Reward Probability and Uncertainty by Dopamine Neurons. Science, 299, 1898-1902.

Findling C, Skvortsova V, Dromnelle R, Palminteri S, & Wyart V (2019). Computational noise in reward-guided learning drives behavioral variability in volatile environments. Nature Neuroscience, 22, 2066-2077.

Forstmann BU, Wagenmakers EJ, Eichele T, Brown S, & Serences JT (2011). Reciprocal relations between cognitive neuroscience and formal cognitive models: opposites attract? Trends in Cognitive Sciences, 15, 272-279.

Fouragnan EF, Chau BKH, Folloni D, Kolling N, Verhagen L, Klein-Flügge M, Tankelevitch L, Papageorgiou GK, Aubry JF, Sallet L, & Rushworth MFS (2019). The macaque anterior cingulate cortex translates counterfactual choice value into actual behavioral change. Nature Neuroscience, 22, 797-808.

Frank MJ & Fossella JA (2011). Neurogenetics and pharmacology of learning, motivation, and cognition. Neuropsychopharmacology, 36, 133-152.

Frank MJ, Doll BB, Oas-Terpstra J, & Moreno F (2009). Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. Nature Neuroscience, 12, 1062-1068.

Frank MJ, Moustafa AA, Haughey HM, Curran T, Hutchison KE (2007). Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. PNAS, 104, 16311-16316.

Fujiwara J, Usui N, Eifuku S, Iijima T, Taira M, Tsutsui KI, & Tobler PN (2018). Ventrolateral Prefrontal Cortex Updates Chosen Value According to Choice Set Size. Journal of Cognitive Neuroscience, 30, 307-318.

Gao X, Gong P, Liu J, Hu J, Li Y, Yu H, Gong X, Xiang Y, Jiang C, Zhou X (2016). COMT Val158Met poly- morphism influences the susceptibility to framing in decision-making:

OFC–amygdala functional connectivity as a mediator. Human Brain Mapping, 37, 1880–1892.

Gershman SJ & Tzovaras BG (2018). Dopaminergic genes are associated with both directed and random exploration. Neuropsychologia, 120, 97-104.

Glimcher PW (2014). Value-Based Decision Making. In Glimcher PW & Fehr E (Editors), Neuroeconomics: Decision making and the brain (2nd edition). London, United Kingdom: Elsevier Academic Press.

Gluth S, Kern N, Kortmann M, & Vitali CL (2020). Value-based attention but not divisive normalization influences decisions with multiple alternatives. Nature Human Behaviour, 4, 634–645.

Gluth S, Spektor MS, & Rieskamp J (2018). Value-based attentional capture affects multi-alternative decision making. eLife, 7, 1-36.

Goodale MA, Milner AD, Jakobson LS, & Carey DP (1991). A neurological dissociation between perceiving objects and grasping them. Nature, 349, 154-156.

Greshake B, Bayer PE, Rausch H, & Reda J (2014). openSNP-A Crowdsourced Web Resource for Personal Genomics. PLoS ONE, 9.

Griffiths TL, Chater N, Norris D, & Pouget A (2012). How the Bayesians Got Their Beliefs (and What Those Beliefs Actually Are): Comment on Bowers and Davis (2012). Psychological Bulletin, 138, 415-422.

Haber SN & Behrens TEJ (2014). The Neural Network Underlying Incentive-Based Learning: Implications for Interpreting Circuit Disruptions in Psychiatric Disorders. Neuron, 83, 1019-1039.

Haber SN & Knutson B (2011). The Reward Circuit: Linking Primate Anatomy and Human Imaging. Neuropsychopharmacology, 35, 4-26.

Haider B, Hausser M, & Carandini M (2013). Inhibition dominates sensory responses in the awake cortex. Nature, 493, 97-100.

Hare TA, Camerer CF, Rangel A (2009). Self-control in decision-making involves modulation of the vmPFC valuation system. Science, 324, 646-648.

Hunt LT, Dolan RJ, & Behrens TEJ (2014). Hierarchical competitions subserving multi-attribute choice. Nature Neuroscience, 17, 1613-1622.

Hunt LT, Kolling N, Soltani A, Woolrich MW, Rushworth MFS, & Behrens TEJ (2012). Mechanisms underlying cortical activity during value-guided choice. Nature Neuroscience, 15, 470-476.

Hunt LT, Malalasekera WMN, de Berker AO, Miranda B, Farmer SF, Behrens TEJ, & Kennerley SW (2018). Triple dissociation of attention and decision computations across prefrontal cortex. Nature Neuroscience, 21, 1471-1481.

Jamali M, Grannan B, Haroush K, Moses ZB, Eskandar EN, Herrington T, Patel S, & Williams ZM (2019). Dorsolateral prefrontal neurons mediate subjective decisions and their variation in humans. Nature Neuroscience, 22, 1010-1020.

James GM, Gryglewski G, Vanicek T, Berroterán-Infante N, Philippe C, Kautzky A, Nic L, Vraka C, Godbersen GM, Unterholzner J, Sigurdardottir HL, Spies M, Seiger R, Kranz GS, Hahn A, Mitterhauser M, Wadsak W, Bauer A, Hacker M, Kasper S, & Lanzenberger R (2018). Parcellation of the Human Cerebral Cortex Based on Molecular Targets in the Serotonin System Quantified by Positron Emission Tomography In vivo. Cerebral Cortex, 29, 372-382.

Jocham G, Hunt LT, Near J, & Behrens TEJ (2012). A mechanism for value- guided choice based on the excitation-inhibition balance in prefrontal cortex. Nature Neuroscience, 15, 960-961.

Juechems K, Balaguer J, Catanon SH, Ruz M, O'Reilly JX, & Summerfield C (2019). A Network for Computing Value Equilibrium in the Human Medial Prefrontal Cortex. Neuron, 101, 1-11.

Kohl C, Spieser L, Forster B, Bestmann S, & Yarrow K (2018). The Neurodynamic Decision Variable in Human Multi-alternative Perceptual Choice. Journal of Cognitive Neuroscience, 31, 262-277.

Körding KP & Wolpert DM (2006). Bayesian decision theory in sensorimotor control. Trends in Cognitive Sciences, 10, 319-326.

Krajbich I, Armel C, & Rangel A (2011). Visual fixations and the computation and comparison of value in simple choice. Nature Neuroscience, 13, 1292-1298.

Kriegeskorte N & Douglas PK (2018). Cognitive computational neuroscience. Nature Neuroscience, 21, 1148-1160.

Lau H & Rosenthal D (2011). Empirical support for higher-order theories of conscious awareness. Trends in Cognitive Sciences, 15, 365-373.

Le Pelley ME, Beesley T, & Griffiths O (2011). Overt attention and predictiveness in human contingency learning. Journal of Experimental Psychology. Animal Behavior Processes, 37, 220–229.

Le Pelley ME, Mitchell CJ, Beesley T, George DN, & Wills AJ (2016). Attention and Associative Learning in Humans: An Integrative Review. Psychological Bulletin, 142, 1111-1140.

Lee D (2013). Decision Making: from neuroscience to psychiatry. Neuron, 78, 233-248.

Lee D, Seo H, & Jung MW (2012). Neural Bases of Reinforcement Learning and Decision Making. Annual Reviews of Neuroscience, 35, 287-308.

Levy DJ & Glimcher PW (2011). Comparing Apples and Oranges: Using Reward-Specific and Reward-General Subjective Value Representation in the Brain. The Journal of Neuroscience, 31, 2011.

Levy DJ & Glimcher PW (2012). The root of all value: a neural common currency for choice. Current Opinion in Neurobiology, 22, 1027-1038.

Leong YC, Radulescu A, Daniel R, DeWoskin V, & Niv Y (2017). Dynamic Interaction between Reinforcement Learning and Attention in Multidimensional Environments. Neuron, 93, 451-463.

Lieder F, Griffiths TL and Hsu M (2018). Over-representation of extreme events in decision-making reflects rational use of cognitive resources. Psychological Review, 125, 1-53.

Liu H, Zakininiaeiz Y, Cosgrove KP, & Morris ED (2017). Toward whole-brain dopamine movies: a critical review of PET imaging of dopamine transmission in the striatum and cortex. Brain Imaging and Behavior, 1-9.

Louie K, Grattan LE, & Glimcher PW (2011). Reward Value-Based Gain Control: Divisive Normalization in Parietal Cortex. The Journal of Neuroscience, 31, 10627-10639.

Ma WJ, Husain M, & Bays PM (2014). Changing concepts of working memory. Nature Neuroscience, 17, 347-356.

Mackintosh NJ (1975). A Theory of Attention: Variations in the Associability of Stimuli with Reinforcement. Psychological Review, 82, 276-298.

Maier SU, Beharelle AR, Polanía R, Ruff CC, & Hare TA (2020). Dissociable mechanisms govern when and how strongly reward attributes affect decisions. Nature Human Behaviour.

Marr D (1982). The Philosophy of the approach. In Vision. New York: WH Freeman. San Francisco.

Martens S & Wyble B (2010). The attentional blink: past, present, and future of a blind spot in perceptual awareness. Neuroscience and Biobehavioral Reviews, 34, 947-957.

Matsumoto M & Hikosaka O (2009). Two types of dopamine neuron distinctly convey positive and negative motivational signals. Nature, 459, 837-841.

Maunsell JHR & Treue S (2006). Feature-based attention in visual cortex. Trends in Neuroscience, 29, 317-322.

Meyer-Lindenberg A, Straub AE, Lipska BK, Verchinski BA, Goldberg T, Callicott JH, Egan MF, Huffaker SS, Mattay VS, Kolachana B, Kleinman JE, & Weinberger DR (2007). Genetic evidence implicating DARPP-32 in human frontostriatal structure, function, and cognition. The Journal of Clinical Investigation, 117, 672-682.

Meyer-Lindenberg A, Kohn PD, Kolachana B, Kippenhan S, McInerney-Leo A, Nussbaum R, Weinberger DR, & Berman KF (2005). Midbrain dopamine and prefrontal function in humans: interaction and modulation by COMT genotype. Nature Neuroscience, 8, 594-596.

Meyniel F & Dehaene S (2017). Brain networks for confidence weighting and hierarchical inference during probabilistic learning. PNAS, 114, E3859-E3868.

Miller EK & Cohen JD (2001). An integrative theory of prefrontal cortex function. Annual Review of Neuroscience, 24, 167-202.

Miller RR, Barnet RV, & Grahame NJ (1995). Assessment of the Rescorla-Wagner Model. Psychological Bulletin, 117, 363-386.

Muller TH, Mars RB, Behrens TE, & O'Reilly JX (2019). Control of entropy in neural models of environmental state. eLife, 8, 1-30.

Nassar MR, Wilson RC, Heasly B, & Gold JI (2010). An Approximately Bayesian Delta-Rule Model Explains the Dynamics of Belief Updating in a Changing Environment. The Journal of Neuroscience, 30, 12366-12378.

Nisbett RE & Wilson TD (1977). Telling More That We Can Know: Verbal Reports on Mental Processes. Psychological Review, 84, 231-259.

Niv Y (Forthcoming). The primacy of behavioral research for understanding the brain.

Niv Y, Daniel R, Geana A, Gershman SJ, Leong YC, Radulescu A, & Wilson RC (2015). Reinforcement Learning in Multidimensional Environments Relies on Attention Mechanisms. The Journal of Neuroscience, 35, 8145-8157.

O'Doherty JP, Cockburn J, & Pauli WM (2017). Learning, Reward, and Decision Making. Annual Review, of Psychology, 68, 19.1-19.28.

O'Hora D, Carey R, Kervick A, Crowley D, & Dabrowski M (2015). Decisions in Motion: Decision Dynamics during Intertemporal Choice reflect Subjective Evaluation of Delayed Rewards. Scientific Reports, 6, 1-17.

O'Reilly JX (2013). Making predictions in a changing world—inference, uncertainty, and learning. Frontiers in Neuroscience, 7, 1-10.

O'Reilly JX & Mars R (2011). Computational neuroimaging: localizing Greek letters? Comment on Forstmann et al. Trends in Cognitive Sciences, 15, 450.

O'Reilly JX & Mars R (2015). Bayesian Models in Cognitive Neuroscience: A Tutorial. In: Forstmann B & Wagenmakers EJ (Editors). An Introduction to Model-Based Cognitive Neuroscience. Springer, New York, NY.

O'Reilly JX, Jbabdi S, & Behrens TE (2012). How can a Bayesian approach inform neuroscience? European Journal of Neuroscience, 35, 1169-1179.

Ott T & Nieder A (2019). Dopamine and Cognitive Control in Prefrontal Cortex. Trends in Cognitive Sciences, 23, 213-233.

Padoa-Schioppa C (2011). Neurobiology of Economic Choice: A Good-Based Model. Annual Review of Neuroscience, 34, 333-359.

Parr T, Rees G, & Friston KJ (2018). Computational Neuropsychology and Bayesian Inference. Frontiers in Human Neuroscience, 12, 1-14.

Passingham RE & Rowe JB (2014). A Short Guide to Brain Imaging: The Neuroscience of Human Cognition. Oxford, United Kingdom: Oxford University Press.

Passingham RE & Wise SP (2012). The Neurobiology of the Prefrontal Cortex: Anatomy, Evolution, and the Origin of Insight. Oxford, United Kingdom: Oxford University Press.

Pearce JM & Hall G (1980). A Model for Pavlovian Learning: Variations in the Effectiveness of Conditioned But Not of Unconditioned Stimuli. Psychological Review, 87, 532-552.

Pearce JM & Mackintosh NJ (2010). Two theories of attention: a review and a possible integration. In Mitchell CJ & Le Pelley ME (Editors), Attention and Associative Learning: From Brain to Behaviour. Oxford, United Kingdom: Oxford University Press.

Pine A, Shiner T, Seymour B, & Dolan RJ (2010). Dopamine, Time, and Impulsivity in Humans. The Journal of Neuroscience, 30, 8888-8896.

Pinker S (2009). How the Mind Works. New York, New York: W. W. Norton & Company.

Persson J & Stenfors C (2018). Superior cognitive goal maintenance in carriers of genetic markers linked to reduced striatal D2 receptor density (C957T and DRD2/ANKK1-TaqlA). PLoS ONE, 12, 1-12.

Polanía R, Woodford M, & Ruff CC (2019). Efficient coding of subjective value. Nature Neuroscience, 22, 134-142.

Preuschoff, K, Hart BM & Einhauser W (2011). Pupil dilation signals surprise: evidence for noradrenaline's role in decision making. Frontiers in Neuroscience, 5, 1-12.

Radulescu A, Niv Y, & Ballard I (2019). Holistic Reinforcement Learning: The Role of Structure and Attention. Trends in Cognitive Science, 23, 278-292.

Rangel A & Clithero JA (2014). The Computation of Stimulus Values in Simple Choice. In Glimcher PW & Fehr E (Editors), Neuroeconomics: Decision making and the brain (2nd edition). London, United Kingdom: Elsevier Academic Press.

Ratcliff R & McKoon G (2008). The diffusion decision model: theory and data for two-choice decision asks. Neural Computation, 20, 873-922.

Redgrave O, Prescott T, & Gurney K (1999). Is the short-latency dopamine response too short to signal reward error? Trends in Neuroscience, 22, 146-151.

Rescorla RA & Wagner AR (1972). A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In Black AH & Prokasy WF (Editors), Classical Conditioning II: Current Research and Theory. New York, United States: Appleton Century Crofts.

Reynolds JNJ, Hyland BI, & Wickens JR (2001). A cellular mechanism of reward-related learning. Nature, 413, 67-70.

Roitman MF, Wheeler RA, Wightman RM, & Carelli RM (2008). Real-time chemical responses in the nucleus accumbens differentiate rewarding and aversive stimuli. Nature Neuroscience, 11, 1376-1377.

Rouault M, Drugowitsch J, & Koechlin E (2019). Prefrontal mechanisms combining rewards and beliefs in human decision-making. Nature Communications, 10, 1-16.

Reutskaja R, Lindner A, Nagel R, Andersen RA, & Camerer CF (2018). Choice overload reduces neural signatures of choice set value in dorsal striatum and anterior cingulate cortex. Nature Human Behaviour, 2, 925–935.

Sarafyazd M & Jazayeri M (2019). Hierarchical reasoning by neural circuits in the frontal cortex. Science, 364, 1-9.

Schultz W, Dayan P, & Montague PR (1997). A neural substrate of prediction and reward. Science, 275, 1593-1599.

Schwarz G (1978). Estimating the dimension of a model. The Annals of Statistics, 6, 461-464.

Seamans JK & Yang CR (2004). The principal features and mechanisms of dopamine modulations in the prefrontal cortex. Progress in Neurobiology, 74, 1-57.

Shen W, Flajolet M, Greengard P, & Surmeier DJ (2008). Dichotomous dopaminergic control of striatal synaptic plasticity. Science, 321, 848-851.

Shenhav A, Straccia MA, Musslick S, Johen JD, & Botvinick MM (2018). Dissociable neural mechanisms track evidence accumulation for selection of attention versus action. Nature Communications, 9, 1-10.

Shiner T, Seymour B, Wunderlich K, Hill C, Bhatia KP, Dayan P, & Dolan RJ (2012). Dopamine and performance in a reinforcement learning task: evidence from Parkinson’s disease. Brain, 135, 1871-1883.

Slifstein M, Kolachana B, Simpson EH, Tabares P, Cheng B, Duvall M, Gordon Frankle W, Weinberger DR, Laruelle M, & Abi-Dargham A (2008). COMT genotype predicts cortical-limbic D1 receptor availability measured with [^{11}C]NNC112 and PET. Molecular Psychiatry, 13, 821-827.

Stahl SM (2013). Stahl's Essential Psychopharmacology: Neuroscientific Basis and Practical Applications, Fourth edition. Cambridge, United Kingdom: Cambridge University Press.

Steinemann NA, O'Connell RG, & Kelly SP (2018). Decisions are expedited through multiple neural adjustments spanning the sensorimotor hierarchy. Nature Communications, 9, 1-12.

Stojić H, Orquin JL, Dayan P, Dolan RJ, Speekenbrink M (2020). Uncertainty in learning, choice, and visual fixation. PNAS, 117, 3291-3300.

Stone JV (2013). Bayes' Rule: A Tutorial Introduction to Bayesian Analysis. United Kingdom: Sebtel Press.

Sutton RS & Barto AG (2018). Reinforcement Learning: an introduction, Second Edition. Cambridge, Massachusetts: The MIT Press.

Ting C, Yu C, Maloney LT, & We S (2015). Neural Mechanisms for Integrating Prior Knowledge and Likelihood in Value-Based Probabilistic Inference. The Journal of Neuroscience, 35, 1792-1805.

Tobler PN, Fiorillo CD, & Schultz W (2005). Adaptive Coding of Reward Value by Dopamine Neurons. Science, 307, 1642-1645.

Tost H, Hakimi S & Meyer-Lindenberg A (2009). Dopamine Dysfunction in Schizophrenia: From Genetic Susceptibility to Cognitive Impairment. In Iversen L, Iversen S, Dunnett S

& Bjorklund A (Editors), Dopamine Handbook. Oxford, United Kingdom: Oxford University Press.

Trudel N, Scholl J, Klein-Flügge MC, Fouragnan E, Tankelevitch L, Wittmann MK, & Rushworth MFS (2020). Polarity of uncertainty representation during exploration and exploitation in ventromedial prefrontal cortex. Nature Human Behaviour.

Tsetsos K, Usher M, & Chater N (2010). Preference Reversal in Multiattribute Choice. Psychological Review, 117, 1275-1293.

van Schouwenburg MR, den Ouden HEM, & Cools R (2010). The Human Basal Ganglia Modulate Frontal-Posterior Connectivity during Attention Shifting. The Journal of Neuroscience, 30, 9910-9918.

Voigt K, Murawski C, Speer S, & Bode S (2019). Hard Decisions Shape the Neural Coding of Preferences. The Journal of Neuroscience, 39, 718-726.

Walton ME, Chau BKH, Kennerley SW (2015). Prioriti sing the relevant information for learning and decision making within orbital and ventromedial prefrontal cortex. Current Opinion in Behavioral Sciences, 1, 78-85.

Wang JX, Kurth-Nelson Z, Kumaran D, Tirumala D, Doyer H, Leibo JZ, Hassabis D, & Botvinick M (2018). Prefrontal cortex as a meta-reinforcement learning system. Nature Neuroscience, 21, 860-868.

Wang XJ (2002). Probabilistic decision making by slow reverberation in cortical circuits. Neuron, 36, 955–968.

Warren CM, Wilson RC, van der Wee NJ, Giltay EJ, van Noorden MS, Cohen JD & Nieuwenhuis S (2017). The effect of atomoxetine on random and directed exploration in humans. PLoS ONE, 12, 1-17.

Watabe-Uchida M, Eshel N, & Uchida N (2017). Neural Circuitry of Reward Prediction Error. Annual Review of Neuroscience, 40, 373-394.

Watanabe E, Kitaoka A, Sakamoto K, Yasugi M, & Tanaka K (2018). Illusory Motion Reproduced by Deep Neural Networks Trained for Prediction. Frontiers in Psychology,

Whitwell RL, Milner AD, & Goodale MA (2014). The two visual systems hypothesis: new challenges and insights from visual form agnosic patient DF. Frontiers in Neurology, 5, 1-8.

Wickens JR, Horvitz JC, Costa RM, & Killcross S (2007). Dopaminergic Mechanisms in Actions and Habits. The Journal of Neuroscience, 27, 8181-8183.

Willems RM, der Haegan LV, Fisher SE, & Francks C (2014). On the other hand: including left-handers in cognitive neuroscience and neurogenetics. Nature Reviews Neuroscience, 15, 193-201.

Wilhelm O, Hildebrandt A, & Oberauer K (2013). What is working memory capacity, and how can we measure it? Frontiers in Psychology, 4, 1-22.

Williams GV & Goldman-Rakic PS (1995). Modulation of memory fields by dopamine D1 receptors in prefrontal cortex. Nature, 376, 572-575.

Wilson RC & Collins AGE (2019). Ten simple rules for the computational modeling of behavioral data. eLife, 8, 1-33.

Wilson RC, Bonawitz E, Costa VB, & Ebitz RB (2020). Balancing exploration and exploitation with information and randomization. Current Opinion in Behavioral Science, 38, 49-56.

EA, Cohen JD (2014). Humans Use Directed and Random Exploration to Solve the Explore–Exploit Dilemma. Journal of Experimental Psychology, 143, 2074-2081.

Wilson RC, Nassar MR, & Gold JI (2010). Bayesian Online Learning of the Hazard Rate in Change-Point Problems. Neural Computation, 22, 2452-2476.

Zajkowski WK, Kossut M, & Wilson RC (2017). A causal role for right frontopolar cortex in directed, but not random, exploration. eLife, 6, 1-18.

Zwaan RA, Pecher D, Paolacci G, Bouwmeester S, Verkoeijen P, Dijkstra K, & Zeelenberg R (2018). Participant Nonnaiveté and the reproducibility of cognitive psychology. Psychonomic Bulletin and Review, 25, 1968–1972.

Appendix A. Recruitment flyer

THE HONG KONG POLYTECHNIC UNIVERSITY
香港理工大學

康復治療科學系
Department of Rehabilitation Sciences

Psychology Experiment:

get paid to play an online game from home!

This experiment is entirely remote:

1. Materials are emailed.
2. Practice and instructions take 10-20 minutes.
3. Full game takes ~15 minutes to complete.
4. Reward can reach up-to HKD200
   (based on performance, paid via the *PayMe* app).

Please note our restrictions:

- No history of neurological impairment.
- Normal or corrected-to-normal vision.
- Be between 18-40 years old.
- Must be right-handed.

If you are interested and fit the criteria, please email me below:

cristian.giron@

For more information about the experiment:

Information Sheet

Appendix B. English and Chinese Task Instructions

Subjects received task instructions by email via links to PDFs saved on Google Drive.

Instructions (in English):

https://drive.google.com/open?id=1N4OJcvc_w8-vktC-T7A6KDIzX00RKGzt

Instructions (in Chinese):

https://drive.google.com/open?id=1z7fXJO4QEUuHIHI00H5tr9PcAAN0XZ03

Appendix C. Printable consent webpage

We need your consent to proceed!

Project Title: Neuromodulators of multi-attribute learning and decision making

The project has been approved by the Human Subjects Ethics Sub-committee (HSESC) (or its Delegate) of The Hong Kong Polytechnic University (Reference Number: HSEARS20190416034).

Investigators Mr Cristian Giron and Dr Bolton Chau
Department of Rehabilitation Sciences,
The Hong Kong Polytechnic University

The purpose of this study. Daily life requires important decisions that are often multifaceted, sometimes demanding us to consider tiny differences between our options. Our brains evolved decision strategies to make good choices, but the use of these strategies varies between individuals. Interestingly, we can predict variations by looking at genetic differences. The purpose of this study is to investigate these decision strategies and analyze your strategy preferences using mathematical models and genetic variation.

Your participation. This session should take you less than 20 minutes. You will be asked to play a computer game that tests your learning and decision making. In this game, you will be shown some pictures and asked to make decisions with button presses. Your responses will be recorded. You will be given opportunities to rest during the experiment.

Benefits. You will earn a cash reward whose amount depends on your performance: automatically earn $50 for finishing the game, *plus* up-to $150 (i.e. totaling $200 depending on score). The findings of this study will also be helpful towards answering a critical scientific question: how do healthy brains make good decisions?

Consent.
If you click the "I consent" button below, you acknowledge the following:

I have read the details of this study above. I voluntarily consent to participate in this study. I understand that I can withdraw from this study at any time without giving reasons, and my withdrawal will not lead to any punishment or prejudice against me. I understand that my personal information will not be disclosed to people who are not related to this study and my name or photograph will not appear on any reports about this study.

I can contact the lead investigator at **cristian.giron@** for any questions about this study. If I have complaints related to the investigator(s), I can contact Ms Vangie Chung, Secretary of the Departmental Research Committee, at (+852) 2766 4329. I know I can print a copy of this consent form.

Print

Do you understand and consent to these terms?

☐ Yes, and I agree to take part in this study.

Important: please do not refresh the page or change tabs after this consent page.

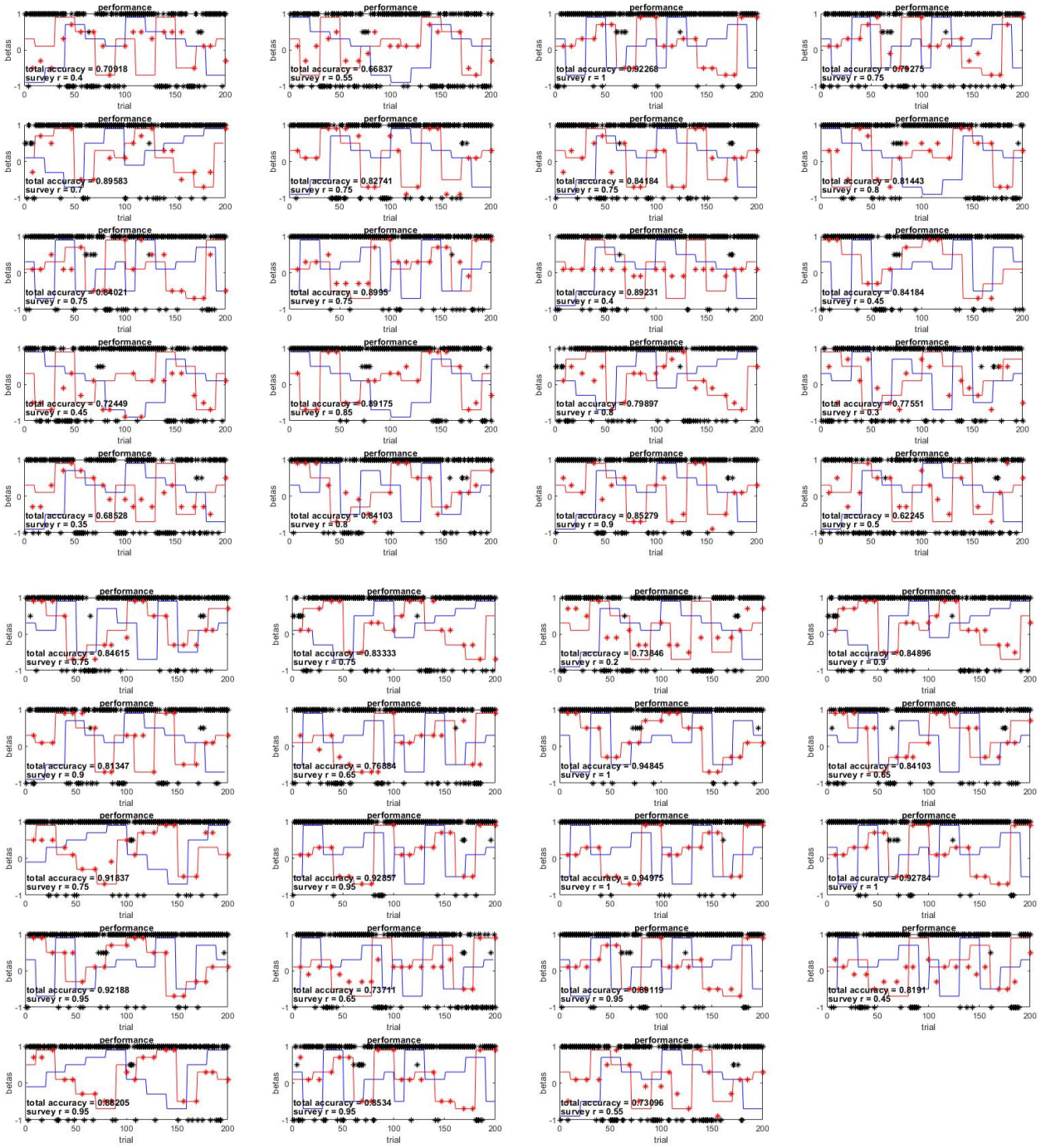
I consent No thanks

Appendix D. Correlation coefficients of task specifications

	LknEV	RknEV	LunEV	RunEV	LknAt	RknAt	LunAt	RunAt
LknEV	1	0.535 ± 0.011	-0.089 ± 0.028	-0.101 ± 0.047	0.191 ± 0.025	-0.054 ± 0.019	-0.097 ± 0.017	0.123 ± 0.023
RknEV	0.535 ± 0.011	1	-0.111 ± 0.043	-0.093 ± 0.025	-0.048 ± 0.02	0.198 ± 0.035	0.123 ± 0.033	-0.1 ± 0.021
LunEV	-0.089 ± 0.028	-0.111 ± 0.043	1	0.532 ± 0.009	-0.087 ± 0.013	0.135 ± 0.025	0.199 ± 0.021	-0.07 ± 0.016
RunEV	-0.101 ± 0.047	-0.093 ± 0.025	0.532 ± 0.009	1	0.169 ± 0.012	-0.107 ± 0.009	-0.079 ± 0.01	0.206 ± 0.018
LknAt	0.191 ± 0.025	-0.048 ± 0.02	-0.087 ± 0.013	0.169 ± 0.012	1	-0.341 ± 0.001	-0.61 ± 0	0.605 ± 0
RknAt	-0.054 ± 0.019	0.198 ± 0.035	0.135 ± 0.025	-0.107 ± 0.009	-0.341 ± 0.001	1	0.606 ± 0	-0.595 ± 0
LunAt	-0.097 ± 0.017	0.123 ± 0.033	0.199 ± 0.021	-0.079 ± 0.01	-0.61 ± 0	0.606 ± 0	1	-0.344 ± 0.001
RunAt	0.123 ± 0.023	-0.1 ± 0.021	-0.07 ± 0.016	0.206 ± 0.018	0.605 ± 0	-0.595 ± 0	-0.344 ± 0.001	1

Shaded colors by quadrant as discussed in **Figure 16**, including standard errors. Left and known expected value, LknEV; right and known expected value, RknEV; left and unknown expected value, LunEV; right and unknown expected value, RunEV; left and known attribute, LknAt; right and known attribute, RknAt; left and unknown attribute, LunAt; right and unknown attribute, RunAt.

Appendix E. Task performance of each subject



The behavioral results for each subject (n = 39). Unknown and known absorbance schedules are indicated by red and blue lines, respectively. Red stars indicate survey response. Black stars at the top and bottom (i.e., on the horizontal line on 1 and 0, respectively) indicate whether subjects

chose the high value treatment on that trial: 1 for yes, 0 for no. Black stars inside each plot indicates that subject's trial options were equivalent. Subject choice and survey accuracy rates are also shown.

Appendix F. Logistic Regression Results

a. Attribute expected values and magnitudes

		int	LknEV	RknEV	RunEV	LunEV	LknAt	RknAt	LunAt	RunAt
First 5 Trials	mean	0.128	-3.189	2.957	-1.133	0.798	-0.442	0.282	-0.363	0.123
	SD	0.390	1.814	1.874	1.668	1.375	0.812	0.952	0.989	1.140
	t	1.804	-9.631	8.641	-3.720	3.180	-2.983	1.624	-2.008	0.593
	df	29	29	29	29	29	29	29	29	29
	p	0.0815569	1.54E-10	1.62E-09	0.0008502	0.0034961	0.0057354	0.1152475	0.0539995	0.557504403
		int	LknEV	RknEV	RunEV	LunEV	LknAt	RknAt	LunAt	RunAt
Second 5 Trials	mean	0.311	-3.707	3.706	-2.405	2.136	0.241	-0.372	0.167	-0.007
	SD	1.679	2.317	2.428	1.747	1.572	1.201	1.181	0.928	1.432
	t	0.830	-7.156	6.826	-6.155	6.075	0.898	-1.409	0.807	-0.023
	df	19	19	19	19	19	19	19	19	19
	p	0.4169938	8.42E-07	1.63E-06	6.48E-06	7.66E-06	0.380387	0.1749662	0.4296244	0.981844031
		int	LknEV	RknEV	RunEV	LunEV	LknAt	RknAt	LunAt	RunAt
Third 5 Trials	mean	0.202	-3.804	3.732	-2.582	2.762	0.172	0.126	-0.417	0.178
	SD	1.105	2.481	2.417	2.144	2.364	1.334	0.831	1.647	1.711
	t	0.879	-7.355	7.406	-5.777	5.604	0.619	0.728	-1.215	0.500
	df	22	22	22	22	22	22	22	22	22
	p	0.3891476	2.32E-07	2.07E-07	8.22E-06	1.24E-05	0.5424105	0.4745444	0.2371182	0.622095654
		int	LknEV	RknEV	RunEV	LunEV	LknAt	RknAt	LunAt	RunAt
Fourth 5 trials	mean	0.038	-4.046	3.792	-3.046	2.522	0.628	-0.163	0.096	0.252
	SD	0.676	2.501	2.470	2.398	1.987	1.683	1.358	0.830	0.978
	t	0.246	-7.051	6.692	-5.536	5.532	1.626	-0.525	0.506	1.123
	df	18	18	18	18	18	18	18	18	18
	p	0.808296	1.41E-06	2.82E-06	2.96E-05	2.98E-05	0.1213635	0.6061908	0.6187876	0.276319857

b. Regressors: Option expected values and informativeness

$$\left(= \frac{unknown\ attribute}{unknown\ attribute + known\ attribute}\right)$$

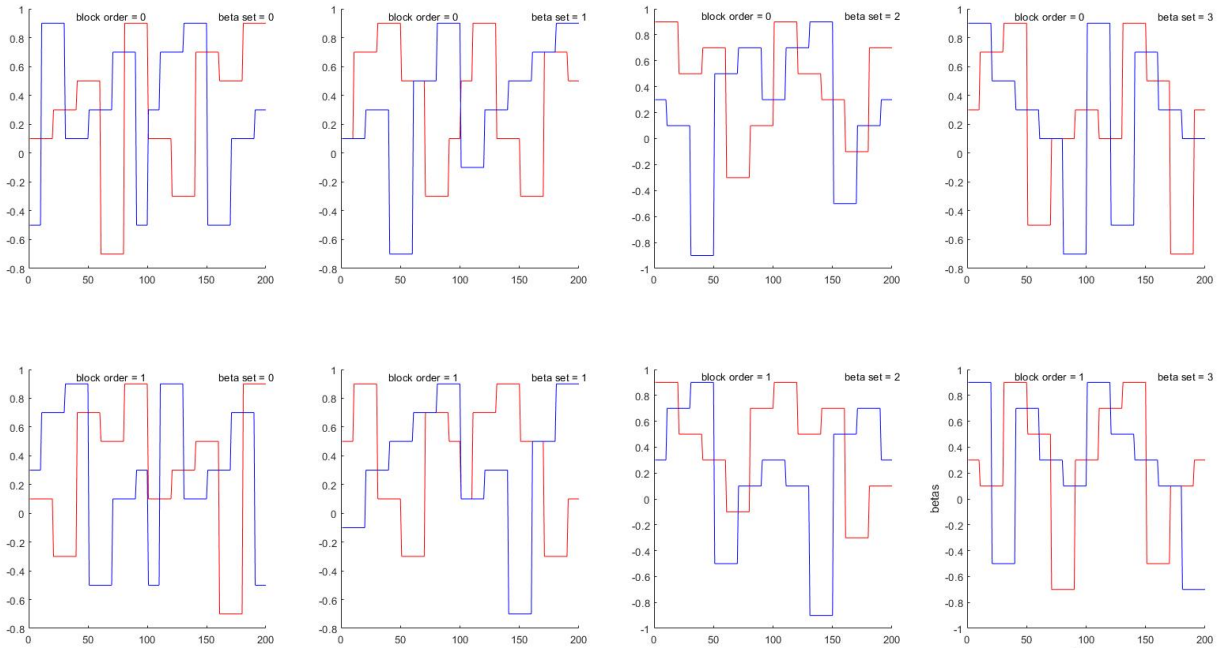
		int	LEV	REV	Linfo	Rinfo
First 5 Trials	'mean'	0.0438698	-2.331075	2.1051865	0.2022397	-0.3020818
	'SD'	0.3674694	1.3625994	1.2142876	0.5311079	0.5886064
	't'	0.7359302	-10.545808	10.687123	2.3473376	-3.1636717
	'df'	37	37	37	37	37
	'p'	0.4664134	1.06E-12	7.27E-13	0.0243679	0.0031097
		int	LEV	REV	Linfo	Rinfo
Second 5 Trials	'mean'	0.2166745	-3.6547919	3.6083526	0.0991934	-0.1355759
	'SD'	0.7884149	2.0592995	2.0538844	0.4623168	0.7454315
	't'	1.505267	-9.7208393	9.6226259	1.1751787	-0.9961745
	'df'	29	29	29	29	29
	'p'	0.1430706	1.25E-10	1.57E-10	0.2494821	0.3274053
		int	LEV	REV	Linfo	Rinfo
Third 5 Trials	'mean'	-0.0104889	-3.4217839	3.4455689	-0.1223169	-0.1597597
	'SD'	0.7849806	2.0087315	1.9849483	0.687017	0.8169946
	't'	-0.0767585	-9.7856044	9.9716889	-1.0227655	-1.1233239
	'df'	32	32	32	32	32
	'p'	0.9392936	3.84E-11	2.43E-11	0.3140889	0.2696571
		int	LEV	REV	Linfo	Rinfo
Fourth 5 trials	'mean'	-0.0230268	-3.9047148	3.9689998	-0.0410684	-0.0547043
	'SD'	0.584877	2.0358955	2.3114119	0.8540865	0.7347735
	't'	-0.2227118	-10.849478	9.7135668	-0.2720075	-0.4211556
	'df'	31	31	31	31	31
	'p'	0.8252207	4.41E-12	6.39E-11	0.7874197	0.6765473

c. Option expected values and informativeness

$$\left(= \frac{known\ attribute}{unknown\ attribute + known\ attribute}\right)$$

		int	LEV	REV	Linfo	Rinfo
First 5 Trials	'mean'	0.0438698	-2.331075	2.1051865	-0.2022397	0.3020818
	'SD'	0.3674694	1.3625994	1.2142876	0.5311079	0.5886064
	't'	0.7359302	-10.545808	10.687123	-2.3473376	3.1636717
	'df'	37	37	37	37	37
	'p'	0.4664134	1.06E-12	7.27E-13	0.0243679	0.0031097
		int	**LEV**	**REV**	**Linfo**	**Rinfo**
Second 5 Trials	'mean'	0.2166745	-3.6547919	3.6083526	-0.0991934	0.1355759
	'SD'	0.7884149	2.0592995	2.0538844	0.4623168	0.7454315
	't'	1.505267	-9.7208393	9.6226259	-1.1751787	0.9961745
	'df'	29	29	29	29	29
	'p'	0.1430706	1.25E-10	1.57E-10	0.2494821	0.3274053
		int	**LEV**	**REV**	**Linfo**	**Rinfo**
Third 5 Trials	'mean'	-0.0104889	-3.4217839	3.4455689	0.1223169	0.1597597
	'SD'	0.7849806	2.0087315	1.9849483	0.687017	0.8169946
	't'	-0.0767585	-9.7856044	9.9716889	1.0227655	1.1233239
	'df'	32	32	32	32	32
	'p'	0.9392936	3.84E-11	2.43E-11	0.3140889	0.2696571
		int	**LEV**	**REV**	**Linfo**	**Rinfo**
Fourth 5 trials	'mean'	-0.0230268	-3.9047148	3.9689998	0.0410684	0.0547043
	'SD'	0.584877	2.0358955	2.3114119	0.8540865	0.7347735
	't'	-0.2227118	-10.849478	9.7135668	0.2720075	0.4211556
	'df'	31	31	31	31	31
	'p'	0.8252207	4.41E-12	6.39E-11	0.7874197	0.6765473

Appendix G. Possible known and unknown DE schedules



Eight possible schedules of the drug effectiveness were developed. Subjects were randomly assigned to one of these schedules in the set – blue lines indicate the schedule of the known DE and red indicates the schedule of the unknown DE.

Appendix H. List of Figures

Appendix I. List of Equations

Page Number	Equation Number	Equation
20	1	$p_x(EV_x) = \frac{\exp\left(\frac{EV_x}{T}\right)}{\sum_{i=1}^{n} \exp\left(\frac{EV_i}{T}\right)}$
23	2	$RPE = \alpha * (received - expected)$
23	3	$V_{t+1} = V_t + RPE$
24	4	$RPE = \alpha * (1 - 0) > 0$
24	5	$V_{t+1} = (0 + (RPE > 0)) > 0$
25	6	$V_{t+1} = V_t + \alpha * (feedback\ value - V_t)$
30	7	$p(hyp\|\ x) \propto p(x\|hyp) * p(hyp)$
30	8	$p(hyp\|x) = \frac{p(x\|hyp) * p(hyp)}{p(x)}$
37	9	Option Value = $\sum$ (β_i * attribute$_i$)
57	10	$TV = [known\ AM * known\ DE] + [unknown\ AM * unknown\ DE]$
66	11	$informativeness = \frac{known\ AM}{known\ AM + unknow\ AM}$
74	12	$informativeness = \frac{unknown\ AM}{unknown\ AM + known\ AM}$
81	13	$\frac{TV - known\ TV}{unknown\ AM} = unknown\ DE$
86	14	$p(\theta \mid x_t) \propto p(x_t \mid \theta) * p(\theta)$
89	15	Θ = (-1, 1)
89	16	$\mathrm{p}(\theta) = \frac{1}{elements\ in\ the\ statespace} = \frac{1}{199}$
90	17	$\theta_{\mathrm{est,t}} = \sum_{i=1}^{199} p_t(\theta_\mathrm{i}) * \theta_\mathrm{i}$
90	18	$Likelihood\ distribution \sim N(x_i, std(\mathrm{p}(\Theta) * \Theta))$
91	19	$p(\Theta) \propto p(\Theta\|\ x_{i-1})$
91	20	$p_{t+1}(\theta_\mathrm{i}) = p_t(\theta_\mathrm{i}\|x_t) * (1 - H) + U(\theta) * H$
93	21	$uncertainty = std([\mathrm{p}(\Theta) * \Theta])$
93	22	$EV = VF + k * IF$
95	23	$EV = [(1 - uncertainty) * VF] + [uncertainty * k * IF]$
95	24	$EV = VF$
96	25	$EV = [VF] + [k * IF]$
96	26	$EV = [(1 - uncertainty) * [VF]] + [(uncertainty) * [k * IF]]$
98	27	$LL = \sum log\ p(EV_{chosen}\|\ parameters, model)$
100	28	$AIC = -2\ x \log(LL) + 2\ x\ numParameters$
100	29	$BIC = -2x \log(LL) + numParameters\ x \log(numObservations)$