Cognitive Architectures, Kinds, and Belief

Joshua Mugg

Forthcoming in *The Nature of Belief.* Edited by Jonathan Jong and Eric Schwitzgebel. Oxford University Press.

Keywords: Natural Kinds, Belief Voluntarism, Wishful Thinking, Acceptance, Reasoning, Dual Process Theory, Inferentially Promiscuous, Settling State, Indirect Measurement

Abstract: To take an empirical approach toward belief, I suggest thinking of belief as a putative kind within the domain of cognitive science. Adopting realist, naturalist, and non-reductionist account of kinds according to which kinds are clusters of causal properties, I argue that a plausible place to begin an inquiry on belief is the architecture of human reasoning. I offer the Sound Board Account of Human Reasoning (S-BAR) (in contrast to Dual Process Theory), according to which there is one reasoning system, which can operate consciously or unconsciously, automatically or controlled, concretely or abstractly, and inductively or deductively. I then argue that, on this architecture, belief will have two properties: inferential promiscuity and settling state. In the final section, I argue that wishful thinkings are beliefs, acceptance and belief are distinct, and argue for a disambiguation of implicit measures.

1. Introduction

'Belief' is often analyzed *a priori* through conceptual analysis: begin with a series of thought experiments and proceed to necessary and sufficient conditions, considering revisions by way of counterexamples until one arrives at reflective equilibrium between definition and intuitions. The goal of this paper is to employ a more empirical approach to the question 'what is a belief?' We can think of belief as a putative kind within the domain of cognitive science. Here I adopt an account of kindhood amendable to the existence of real kinds in the cognitive domain, outline an empirically motivated account of the cognitive architecture of human reasoning, and draw out some key properties of belief on that cognitive architecture. I then use those properties to argue for a form of direct belief voluntarism and distinguish belief from some belief-like states. Since the properties of belief are drawn from a cognitive architecture that is testable, the metaphysics of belief I offer is indirectly empirically testable. I conclude by sketching some implications for indirect measurement in psychology.

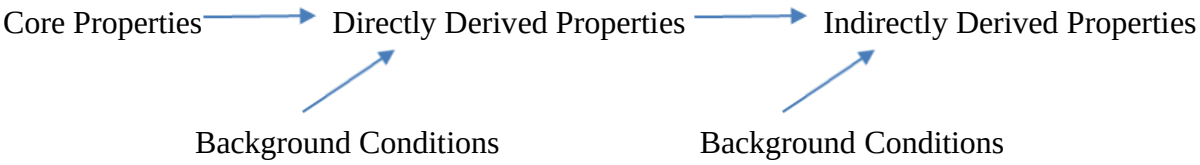
2. Cognitive Kinds

A major obstacle to integrating work on the nature of belief with empirical work is the question of how to operationalize 'belief' for psychological experimentation, given that any operational definition will be controversial when considered against the various accounts on offer from the metaphysics of belief literature. I suggest that we think of belief as a putative kind within the domain of cognitive science. Causal structure seems to be central to kindhood rather than mere property co-instantiation, since science is particularly interested in distinguishing causation from mere correlation, and inferences involving kinds are grounded in causality. Most naturalistic theories of kindhood agree. For example, Kornblith explains that "it is precisely because the world has the causal structure required for the existence of natural kinds that inductive knowledge is even possible" (1993, 35), and Khalidi "assume[s] that uniformities in nature are due to regular and

stable connections between causes and effects, and that these *causal relations* are the metaphysical bases of scientific induction and epistemic practices" (2023: 5). Here I adopt a simple causal theory of kinds, according to which "certain properties or conjunctions of properties that are causally connected with others in systematic ways can be considered natural kinds" (Khalidi 2023: 5, c.f. Craver 2009). While this view is compatible with multiple accounts of explanation, it fits most naturally with a view that prioritizes causal explanations.

Khalidi has offered a more fleshed out version of the simple causal theory of kinds, according to which kinds are clusters of causal properties, blurring the line between properties and kinds. Indeed, on this account, kinds are property clusters. Put slightly differently, we can think of kinds as "highly connected nodes in causal networks" that can be represented as highly connected vertices in directed causal graphs (Khalidi 2018: 1389).[1] Consider the following, amended from Khalidi (2018: 1387):

Figure 1:



On this account, the core properties are the kind, which, with background and other properties, cause derived properties directly or indirectly. While this figure gives the impression that causation is simple and linear, Khalidi grants that "in many cases we observe a web-like network of such causal relationships whereby some causal processes interact with other causal processes (e.g. when gold melts, the properties of the chemical bonds of solid gold interact with the properties of temperature and atmospheric pressure)" (2018: 1387).[2] If there is a node in a causal network, that node will have all the markings of a kind—figuring in explanations, predictions, and inductions. Thus, the node will meet the empirical adequacy we need from an account of natural kinds because of the causal connectedness of the node. Note that none of this requires a specific mechanism (contra Boyd's (1999) Homeostatic Property Cluster account of kinds).[3] This account is *realist* insofar as causal structures are real. While human interests might drive which causal structures we pick out in some domain of inquiry, this does not result in a subjectivism about kinds since such cases merely involving attending to distinct parts of the underlying causal structure of reality. The simple causal theory is *naturalistic* in that it takes kinds to be discovered by empirical investigation rather than though a purely conceptual analysis. As such, the naturalism this account adopts is

[1] Following Khalidi, I will speak of 'kinds' instead of 'natural kinds.' 'Natural kinds' are sometimes contrasted with 'artificial kinds.' However, some artificial kinds can properly be called natural, such as synthetically produced compounds. Sometimes the 'natural' in 'natural kinds' is supposed to contrast with kinds in special sciences. I am skeptical that these domains can be distinguished so neatly, and within the cognitive science there are plausibly kinds that that straddle biology, psychology, and sociology.

[2] Khalidi appeals to Woodward's (2003) influential interventionist account of causation.

[3] At times, Boyd seems to loosen his requirement that there *must* be some singular mechanism holding properties in homeostasis, such as when he says, "either the presence of some of the properties in [the cluster] *F* tends (under appropriate conditions) to favor the presence of the others, or there are underlying mechanisms or processes which tend to maintain the presence of the properties in [the cluster] *F*, or both" (Boyd 1989, 16). If one drops the mechanism requirement, the result is a simple causal theory of kinds.

more methodological than metaphysical. It is compatible with *non-reductivism* because there is no requirement that the causal relations be reduced to some lower level. Indeed, the account is tailor-made for the existence of inter-domain kinds—kinds that traverse so-called 'levels of reality' insofar as causal thickets travers layers of reality (Wimsatt 1994). Furthermore, because kinds are 'highly connected' nodes, there will be border-cases, both for whether something rises to the level of a kind and whether some particular entity counts as a member of a kind. This is so because this account of kinds is *non-essentialist.*

Khalidi's account has recently been applied to several putative cognitive kinds such as working memory (Gomez-Lavin 2021), memory (Colaço 2022), cognitive bias/heuristic (Stanovich, Toplak, and West 2020, Khalidi and Mugg 2023), and concepts (Khalidi 2017). My aim here is to extend this application to belief. On this account, if belief is a kind, then it is a node in a causal network—a cluster of causal properties. If cognitive kinds are individuated in terms of their causal role, a promising approach to distinguishing belief from acceptance, credence, propositional faith, and imagination—what I will call our doxastic taxonomy—is to do so with reference to a causal model cognitive architecture.[4] As such, my method is to start with a cognitive architecture of human reasoning developed to explain empirical data from cognitive and social psychology, and then consider the empirical adequacy of belief and doxastic-like states within that cognitive architecture. Of course, the doxastic taxonomy will be a part of the cognitive architecture since each kind is a node in the causal network, and so will itself figure in explanations of empirical data. The taxonomy will do so, however, via cognitive architecture which outlines the role each kind plays. We use experimental evidence to build, argue for by way of inference to the best explanation, and test cognitive architectures, and views of the relation of belief and doxastic-like attitudes will fit better or worse with various cognitive architectures. Those who find particular views on, say, the relation between belief and credence for reasons other than cognitive architecture may use my theoretical work here in an inference to the best explanation for one view of cognitive architecture over another.

The first thing to note is that, even though 'belief' comes to us from folk psychology, psychologists rely on 'belief.' For example, cognitive psychologists use 'belief' in predictions and explanations in confirmation bias, rebound effect, false-belief task, belief-bias in syllogistic reasoning, and placebo effects.[5] In cognitive science of religion, anthropologists, psychologists, and neuroscientists take religious *belief* (among other things) as explanadum. Belief ascriptions play important roles in normal-functioning adults and many developmental psychologists are interested in when humans gain the ability to attribute beliefs (especially false-beliefs) to others. While social psychologists tend to emphasize *attitudes,* which may or may not be belief-involving, they frequently rely on 'belief' as a contrast to what it is they study in implicit bias, especially those focused on implicit biases that are opposed to subjects' professed values (e.g. Greenwald & Banaji 1995), and they employee belief language in studies and measurements of political polarization (e.g. Rollwage et. al 2019, Inbar and Lammers 2012). This is *prima facia* evidence that belief is highly connected within the causal network that psychology aims to map. What is underdetermined is whether belief is a single node or a hodge-podge of nodes that we should distinguish. If 'belief' as currently deployed by psychologists does refer to a host of nodes, getting clear on cognitive architecture can help in providing a better taxonomy.

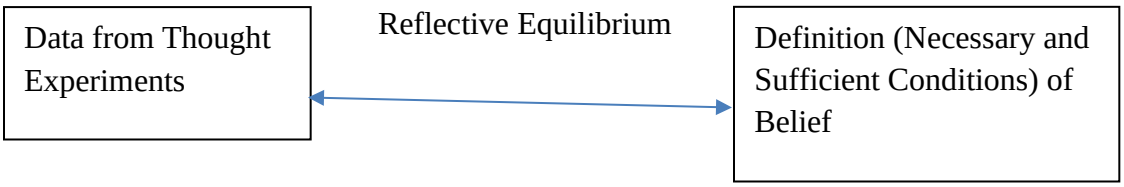
[4] Beliefs are states, but not all kinds need be states, as kinds might be processes (e.g. learning), objects (e.g. neuron), or capacities (e.g. mindreading). I will remain agnostic on which ontological category grounds which.

[5] See Porot and Mandelbaum (2020) for a survey of the prevalence of 'belief' in contemporary cognitive science, especially in placebo effect, attribution theory, theory of mind, and comparative psychology.

Two complications require comment. The first is the existence of other folk psychological belief-like terms, such as 'acceptance,' 'faith,' 'imagining,' etc. Again, these putative kinds are often analyzed *a priori* through conceptual analysis. In my view, a difficulty with such a purely conceptual analysis method of distinguishing these terms is that different philosophers settle the reflective equilibrium in different (incompatible) ways. For example, consider someone confiding to a close friend the following about their partner: "I do not know what to believe, whether she'll stay with me or not; I'm in doubt about the matter. Even so, I have faith that she'll stay" (Howard-Snyder 2019: 116). While non-doxasticist Howard-Snyder thinks such cases show that faith need not involve belief because we should take such talk literally, doxasticists, such as Malcomb and Smith (2017), have the competing intuition that we should think of such speech as 'loose.' It seems that the addition of cases is unlikely to break the disagreement. However, this does not imply that conceptual analysis is useless. While conceptual analysis may be unlikely to deliver an empirically fruitful doxastsic taxonomy *on its own*, it can offer putative causal properties associated with these putative kinds. Conceptual analysis gives us a starting point when considering how a cognitive architecture should divide putative doxastic kinds.

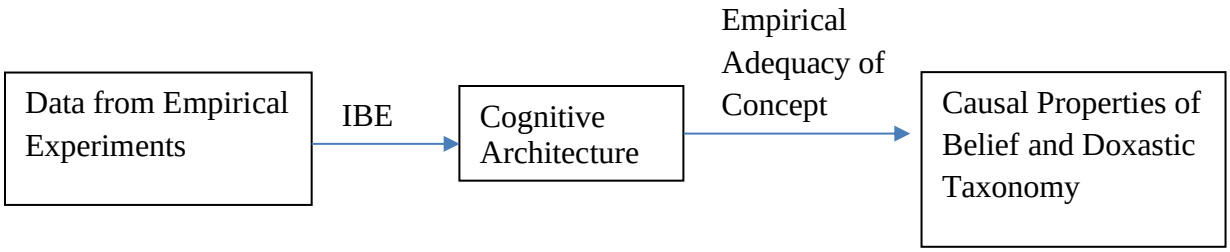
An additional complicating factor in the cognitive science of belief is that, while beliefs are sometimes posited by cognitive scientists, they also sometimes serve as explanadum (e.g. how does belief attribution work? Or how do religious beliefs arise?). If belief is to be explained by the working of the cognitive system, that does not imply that belief itself is part of the workings of the cognitive system. It is possible that folk psychological belief attribution might be predictively fecund without being reducible to actual cognitive underpinnings, as Curry argues (2021a: 1077). Curry (2021b) draws a distinction between *attitudes of* belief—that which the folk attribute to one another—and *cognitive states of* belief—that which cognitive scientists study. If Curry is correct, there would be two distinct putative nodes: one for folk attribution and one for scientific attribution. If one thinks that cognitive scientists, themselves being human, are just engaged in a more rigorous method of folk psychological attribution, then there need not be this bifurcation. While I am more sympathetic to this latter viewpoint, I will not argue for it here.[6] Instead, I will simply note that if Curry is right, then this is a paper about the cognitive states of belief rather than attitudes of belief.

Figure 2: Purely Conceptual Analysis



[6] Briefly, cognitive scientists seem interested in the things that the folk attribute to one another. That is why we (the folk) care about what the cognitive scientists are studying. As such, in my view, psychology and folk psychology cannot be divorce in the way that physics and folks physics can.

Figure 3: Empirical Analysis



I will consider two objections to my method. First, one might object that belief is normative in that it is essentially regulated by truth (e.g. Shah and Velleman 2005), and therefore cannot be studied by science which is purely descriptive. However, there are many scientific inquiries that invoke norms of various kinds (e.g. *adaptiveness* in biology, *prejudice* in social sciences, *bias* in cognitive and social psychology). While there are various philosophical proposals to "naturalize" normativity, none of them need to succeed for my method. Empirical work on reasoning and decision-making is shot-through with normative rules and principles. The heuristics and biases program studies how it is that humans systematically deviate from the norms of reasoning. Thus, this research program crucially relies on norms. Indeed, 'bias' and 'heuristic'—phenomena that cognitive scientists research—are identified by and defined, in part, by appeal to norms. For example, in their introduction to a seminal collection of papers, psychologists Gilovich and Griffin define biases as "departures from the *normative rational theory* that served as markers or signatures of the underlying heuristics" (2002, 3, emphasis mine). Similarly, Stanovich, West, and Toplak write: "The term 'biases' refers to the systematic *errors* that people make in choosing actions and in estimating probabilities, and the term 'heuristic' refers to why people often make these errors—because they use mental shortcuts (heuristics) to solve many problems" (2016, 1110; emphasis mine). Kahneman says of participants in some of his experiments that "[t]he reliance on the heuristic caused predictable biases (systematic *errors*) in their predictions" (2011, 7, emphasis mine). Finally, even though he has a rather different position on the nature and prevalence of bias, Gigerenzer states: "A bias is a systematic discrepancy between the (average) judgment of a person or a group and a *true value or norm*" (2018, 306, emphasis mine). Thus, there seems to be broad consensus in cognitive science that biases are systematic departures from norms of rationality. This does not automatically imply that heuristics and biases are cognitive kinds. It could turn out that 'heuristics' are a grab-bag of rules that bear little resemblance to one another.[7] My point is that the cognitive scientists studying heuristics and biases are more than happy to appeal to norms of rationality in empirically investigating the cognitive architecture of the mind, especially at the outset of an inquiry when they are deciding which of the many phenomena they will investigate. Insofar as we are willing to permit a scientific study of heuristics and biases, treating them as putative cognitive kinds or at least as phenomenon to be studied scientifically, we should be willing to treat belief in the same way. Thus, if belief is inherently normative, that does not count against treating it as a putative kind.

Second, one might worry about wedding doxastic taxonomy to some specific controversial cognitive architecture, since we must then wait on a completed cognitive architecture to determine the kindhood of the doxastic-like states. After all, we must appeal to doxastic-like states in

[7] Khalidi and Mugg (2023) argue that 'heuristics' are not a kind for precisely this reason, but argue that some specific heuristics likely are kinds. They argue that myside bias is in fact a heuristic that is a kind.

outlining the explanandum for which cognitive architectures are built as explanations. While this method may result in a temporary impasse, it need not. With taxonomies of doxastic-like states from the various cognitive architectures in hand, we can compare similarities and differences. One possibility is that the various cognitive architectures, though differing in details of the overall cognitive processing, will produce similar taxonomies, indicating that some doxastic-like states cluster to constitute cognitive kinds. In that case, we need not first determine which cognitive architecture is correct in order to determine whether some of the doxastic-like states cluster into kinds. That would be a happy result for the inquiry into doxastic taxonomy. Of course, it may turn out that the taxonomies do not match up very well, in which case we would need to wait on which cognitive architecture wins out. In my view, this isn't problematic, since it ties competing doxastic taxonomies to competing empirical theories, which should be in principle testable. Thus, we would be able to indirectly test competing doxastic taxonomies, and empirical traction on doxastic taxonomies would be progress. Consequently, tying belief to cognitive architecture is methodologically appealing.

3. Cognitive Architecture

A plausible cognitive architecture to begin with might be dual-process theory (DPT), since some philosophers working within DPT have already deployed a similar method to the one I recommend, moving from cognitive architecture to an account of the nature of belief. On DPT, human reasoning is divided between two types of processing, one fast, automatic, and autonomous (i.e. not involving working-memory) (type-1), the other slow, controlled, and involving working-memory (type-2).[8] Dual-system theory (DST) accepts the type-1/type-2 distinction and adds that each type of processing is underwritten a distinct system: System 1 and System 2.[9] Each system might utilize their own kinds within the doxastic taxonomy. For example, Frankish (2004) distinguishes belief (or belief$_1$) from superbelief (or belief$_2$) by appeal to System 1 and System 2.[10] On his account, 'belief$_1$', which is the belief-state type of System 1, is involuntary and behavioral-dispositional. As such, a subject's beliefs$_1$ will always be coherent and ideally rational. Belief$_2$, which is the belief-state of System 2, is voluntary, and representational.[11]

While I agree with Frankish's method of beginning with a cognitive architecture and then moving to doxastic taxonomy, there are two worries. First, his doxastic taxonomy locates irrationality *entirely* within System 2, whereas empirically motivated accounts of DPT/DST locate irrationality (largely) in System 1. As a result, his version of DST is non-standard and has not been empirically motivated. Second, there are several reasons to be skeptical of DPT/DST, at least for reasoning, judgement, and decision-making. There are worries about the extent to which the properties that divide type-1 and type-2 processing cross-cut one another (see Keren & Schul 2009, Kruglanski & Gigerenzer 2011), the *ad hoc* nature of adopting various property distinctions in distinct explanations (Keren 2013), and that many of the properties distinguishing the processes admit of degrees (Kruglanski 2013, Gigerenzer 2010). In response to these worries, Evans and

[8] Samuels (2009) applies Boyd's account of kinds to DPT, treating type-1/type-2 as putative cognitive kinds, which has largely become the standard way of understanding DPT in the literature.

[9] System 1 is typically thought of as a collection of module-like systems, while System 2 is generally taken to be a unified system separate from System 1 (e.g. Evans 2011, Khaneman 2011, Stanovich 2011) or a virtual system realized in the interaction between the collection of module-like systems (e.g. Carruthers 2009 and Frankish 2004). See Samuels (2009) who first drew attention to this distinction.

[10] Gendler (2008a, 2008b) similarly draws on the dual-process tradition in distinguishing belief, reason, and desire from *sui generous* new kinds 'alief', 'eason', and 'cesire.' See Mandelbaum (2013) for criticism.

[11] Frankish says that belief$_2$/superbelief is Cohen's (1992) 'acceptance.'

Stanovich (2013) agreed that the various properties distinguishing the types of processes do cross-cut, that both types of reasoning can be normatively correct, and so they adopt working memory involving/not-involving as the singular property that distinguishes the two processes. However, this trivializes DPT and, consequently, severely undercuts its explanatory and predictive power (Mugg 2016a). While a cursory glance at the social psychology literature may seem to support dual-process theory, given its emphasis on implicit measurement as distinguished from explicit measurement, a closer examination reveals we should not equate 'implicit' with 'type-1' nor 'explicit' with 'type-2.' For example, in their review of implicit measures in social psychology, Gawronski and De Houwer (2014) question the assumption that "indirect measurement procedures provide a window into unconscious [or evolutionarily old] representations, whereas direct self-report measures reflect conscious representations (e.g. Greenwald & Banaji 1995)" (297). On their view, this is an empirical claim that must be tested. One way to do so is to show that the implicit and explicit measurements do not correlate well with one another, indicating that they measure distinct attitudes. However, Grawonski and De Houwer direct attention to studies which find that implicit and explicit measurements do correlate. While there are studies that find that implicit and explicit measurement do not correlate, they argue that this may be due to the low reliability of many implicit measurements. As such, they warn against thinking that implicit measures "provide direct access to mental associations that are activated automatically upon the encountering a relevant stimulus," going on to explain that "responses on indirect measurements procedures are the product of multiple distinct processes that jointly influence performance on the task" (2014: 301). They also challenge the claim that implicit measurements tap evolutionarily old representations that are hard to change while explicit measurements tap evolutionarily new representations that are more easily altered. While they cite some research that supports such mappings, they point out that there is a "large body of research showing the opposite pattern (e.g. Gawronski and LeBel 2008, Olson and Fazio 2006)" (2014: 298). Grawonski and De Houwer suggest that researchers would do well to "focus on underlying mechanisms with regard to the measures themselves as well as their capability to predict behavior" (2014: 303, see also Nosek, Hawkins, and Frazier 2011). Doing so requires getting specific about cognitive architecture, and, given the criticism of the received version of DPT, it will be helpful to have an alternative that we may apply to our present inquiry.

I will focus on my Sound Board Account of Human Reasoning (S-BAR) (Mugg 2018), according to which there is one reasoning system, which can operate consciously or unconsciously, automatically or controlled, concretely or abstractly (i.e. moving from content-laden argument to attending directly to form of argument), and inductively or deductively. I deny that the properties previously used to distinguish type-1 from type-2 processing cluster. Thus, I am not claiming that there is one reasoning system that operates in a type-1 way sometimes and a type-2 way at other times. On this account, for any reasoning process and any property pair, the reasoning system will operate in a definite way; it will not, for example, operate in a working-memory involving and non-involving way at the same time.[12] After the reasoning system outputs a representation, it can readjust and rework the reasoning process, just as a sound-mixing board operates, and can recruit additional representational states from memory, adding such representations just as a sound-mixer might add a baseline to a track. Since modes of operation do not cluster, that some of the properties

[12] This makes the account is in principle testable: should there be evidence of contradictory beliefs that arise from simultaneously operating reasoning processes, this would be strong evidence against my account. Sloman (1996) argues that humans possess contradictory beliefs, but Mugg (2013, 2018) as argued that there is not good evidence that these contradictory beliefs arise from simultaneously operating reasoning processes.

(e.g. fast/slow) admit of gradation is not a problem. Consider the familiar/unfamiliar distinction. While on DPT, type-2 processing will be required for unfamiliar cases (see esp. Greene 2014), familiar/unfamiliar is clearly a matter of gradation, and the extent to which a problem is novel depends on how much quickly and easily the subject is able to adjust the reasoning system into the proper mode (Dale 2020).

The initial settings of the reasoning system are not set by the system itself, since otherwise the reasoning system would first have to determine how to set itself, which is itself a reasoning process, thereby leading to a vicious regress. How it is set depends on external factors (e.g. wording of a question or problem) and internal factors (e.g. available mindware (Stanovich 2011), availability of working-memory). Whether a subject simply accepts the resulting representation and ceases the reasoning process or (effortfully) feeds the representation back through the reasoning system also depends on both internal and external factors, such as thinking dispositions and available mindware in the former and time allowance for the latter, in addition to the motivation an individual has for 1) getting the answer correct and 2) non-epistemic factors such as whether the belief would be beneficial, comforting, disconcerting, etc.

A number of philosophers have argued that we need accounts of control as graded, both in action and cognition. [13] S-BAR allows for the automatic-controlled distinction to be a matter of degree rather than kind. We can think of a process as more controlled the more it has three features: 1) the more times a subject reruns a process, 2) the more modes of operation are altered in the re-running of the process, and 3) the more doxastic states or heuristics that are recruited. The more a process has these feature, the more controlled it is. Conversely, subjects who accept the initial output of the reasoning system are engaged in a process with little or no control. Control, on this account, is not a slide or switch on the reasoning system, but (weakly) emerges from whether and how the process is rerun.[14]

It will be helpful to see how this account of human reasoning explains some of the findings from the heuristics and biases literature. Belief bias is the phenomenon in which the believed truth of a conclusion (and sometimes premises) of an argument interfere with the judgment of the validity of the argument. While the reasoning system's concrete mode is useful in determining whether a proposition is true, especially when recruiting other beliefs, it is not helpful in determining whether the argument is valid, since all that is relevant to an argument's validity is the form of the propositions in question. Consider the following two arguments from Evans et al. (1983):

Argument 1:

1. No cigarettes are inexpensive.
2. Some addictive things are inexpensive.
3. Therefore, some addictive things are not cigarettes.

[13] For example, Wu (2016) defines control and automaticity in a way that allows for more or less control of an action relative to various features. On his account, a *feature* of an action is controlled just in case the action has that feature because of the subject's intention to act in that featured way. Automaticity is defined as not controlled. This allows for an action to have more or less controlled features. Jennings goes further, arguing that to account for skilled behavior, we must distinguish between two kinds of control: flexible control and reliable control, which exist along distinct and opposing continuums (2020: 172).

[14] One might wonder whether mode alteration, process rerunning, and doxastic recruitment really should be group together under the heading of control. Here I group them together because these are the features of altering initial responses on S-BAR.

Argument 2:

4. No addictive things are inexpensive.
5. Some cigarettes are inexpensive.
6. Therefore, some addictive things are not cigarettes.

Subjects overwhelming accepted both as valid (92% in both cases), even though only the first is valid. In reasoning concretely, subjects are attending to the truth of each proposition. According to S-BAR, correctly judging Argument 2 as invalid requires switching modes. For example, one could switch into an abstract mode, in which the subject attends solely to the form of the propositions:

4. No A are I.
5. Some C are I.
6. Therefore, some A are not C.

Abstraction is the process of decoupling certain parts of the problem from others. In this case, decoupling the form of the propositions in question from their believed truth-value, but decoupling will occur in other reasoning processes as well (e.g. an individual's race from their qualifications, consequences of an action from intent of an action, and raw probability from probability given priors).

One might object to the soundboard metaphor on grounds that there is someone or something adopting the output or rerunning the process, which raises worries about a homunculus fallacy. Applied to this case, the worry is that there is a little homunculus in the head that has a reasoning-soundboard. She listens to the output of the reasoning-soundboard and then *decides* whether to adopt or rerun, and which modes to alter or doxastic states to recruit. But now I have just pushed the account back because we may ask on what basis the homunculus decides. There is still "that conveniently intelligent but opaque agent…lurking…under the guise of a central executive or supervisory attentional system assumed to direct processes that are not 'automatic' " (Monsell and Driver 2000: 3). I have two responses. First, positing a homunculus is not problematic when the homunculus is simpler than the subject—when the homunculus does not do the very thing that is supposed to be explained. In the case of rerunning the process or accepting the output, the homunculus has only two options: accept or rerun. There is no *a priori* reason to think that mode alteration and doxastic recruitment are carried out by the same subsystem. Thus, insofar as I am positing homunculi, they are not of the worrying kind. Secondly, I take these questions to be fruitful ground for further empirical exploration of S-BAR: under what conditions do subjects rerun reasoning processes? Under which conditions do subjects alter modes? Which modes? Empirically founded answers will likely be complex, with variance across individuals and contexts. Such particularity is not worrisome, since my goal is to flesh out the causal nexus and identify nodes, rather than find exceptionless laws.

4. Properties of Belief

Conceptual analyses of belief offer a plethora of properties associated with belief. They are said to be semantically evaluable, acquired through perception and reasoning, difference-makers to behavior, sensitive to evidence, and give rise to Moore's Paradox. While S-BAR is compatible with these properties, my method is to examine which properties fall out of S-BAR—the properties associated with belief from conceptual analyses are mere suggestions. Here I will examine two

properties of belief. On S-BAR, the reasoning system takes beliefs (*inter alia*) as input.[15] I will argue that the way this happens on S-BAR implies that belief is inferentially promiscuous. Furthermore, on S-BAR, sometimes outputs are fed back through the reasoning system. This may happen multiple times. The resulting representational state of the reasoning process is a belief. Put differently, it is the job of the reasoning system to produce and modify beliefs, and so the resulting representational state of a reasoning process is a belief. I will argue that this entails that belief is, in some sense, a settling state.

Beliefs interact with beliefs and other cognitive states in reasoning processes. I am in partial agreement with Mandelbuam here, who defines inferential promiscuity as "*able to* serve as premises in inference" (2014: 30, emphasis mine). That is a fairly low bar since it only amounts to the *ability* of a belief to serve as a premise in an inference. I take it that a belief will *tend to* serve as a premise in inferences to which that belief is relevant. This is so because beliefs are recruited for reasoning processes, often automatically. The opposite of 'inferentially promiscuous' is 'cognitively quarantined,' which is familiar in accounts of pretense (Leslie 1987, Nichols and Stitch 2003). A cognitively quarantined representation will not interact with very many other beliefs, and perhaps not interact with any beliefs at all. A paradigmatic example of a cognitively quarantined cognitive state is an imagining. An imagining's being cognitively quarantined explains why it is possible for a child to wield a stick as though it is a sword while recognizing simultaneously that the stick-sword will not cut him.

Because the final output of a reasoning process is a belief, on this account, belief is a 'settling state' in the sense that when one believes that p, one has settled for themselves whether or not p. Just as we can think of "intending" as settling the question of what one will do, so "belief" settles the question of what is the case (Hieronymi 2006, 2008). Here I am in agreement with usage in recent debates on the relation between faith and belief, where belief and disbelief are taken to be 'settling states' in the sense that when one (dis)believes p, one has settled whether or not p. Importantly, this is common ground for the doxasticist and non-doxasticist about faith (i.e. those who think that belief is partly constitutive of faith and those who do not. See Howard-Snyder (2013) and McKaughan (2013) for the former, Rettler (2018) and Mugg (2022) for the latter). The representations that are fed back into the reasoning system will carry some of the same properties often associated with beliefs listed above. For example, it will be evaluable (perhaps semantically) since there is an evaluation of whether to accept the representation or rerun the process, and an accepted representation will tend to enter into inferences with other beliefs. However, a representation that is immediately fed back into the system will not have effects on thought and behavior in the same way as the final representation.[16]

5. Towards a Doxastic Taxonomy

It will not be possible to explain how belief differs from every belief-like state in this short chapter. Instead, I will begin by examining the version of belief voluntarism that S-BAR supports,

[15] In addition to *beliefs*, the reasoning system can accept claims for the sake of argument in a line of reasoning, engage in arguing by *reductio ad absurdum*, etc. Such inputs are not likely beliefs, since their dispositional profile will look more like an imagining than a belief (more on this below).

[16] Spinozans argue that accepting a proposition is automatic while rejecting it is effortful, and that automatically accepted propositions do affect thought and behavior (Gilbert 1991, Levy and Mandelbaum 2014, Mandelbaum 2014). Space does not permit me to enter into detailed dialogue with Spinozan accounts here. Instead, let me register that Spinozans claim that the subsequently rejected representation affects thought and behavior only momentarily. Namely, they affect thought and behavior until they are rejected. As such, my claim that initial outputs do not affect thought and behavior *in the same way* as final outputs is not contradicted by Spinozan theories.

since belief-involuntarism is often used as a main way to distinguish belief from belief-like states such as acceptance, hypothesis, propositional faith, or trust.

5.1 Belief Voluntarism and Wishful Thinking

If the reasoning system is operating in a low working-memory and fast mode, then the subject will have a response to the question quickly. The subject may then rerun the reasoning process, which is a voluntary process, to some extent. For example, if the stakes are high enough, the subject may effortfully double-check their response. Choosing to double-check is (often) a voluntary act as much as voluntarily kicking a ball. When one double-checks their answer, they generally hope to ensure that they were right, but that is not always the case. When I double check my estimate that I will be late to an important appointment, I hope that I will get a different answer. Still in both cases, the choice to rerun is voluntary, and a belief is the result.

Such a suggestion does not imply that one can believe *anything* at will in the sense that one can raise their hand at will, and this is so because despite my ability to continue the reasoning process, I cannot *guarantee* a different outcome. This is not troubling for belief voluntarists. Conceptually, it does not follow from the fact that the decision to rerun the reasoning process did not result in the desired belief that, counterfactually, if the desired belief had resulted it would not have been voluntarily formed. We often attempt to do things and fail. These attempts are voluntary even if we sometimes fail to achieve what we set out to achieve. I can voluntarily kick a ball hoping to make it to first base, and if I achieve kicking the ball and making it to first base, my action was surely voluntary. But if I kick the ball in the foul zone or it is caught, it is not as though my action loses its voluntariness. The point holds even in classic examples of a voluntary action one can engage in 'at will,' such as raising one's hand. Suppose someone holds down my hand just as I start to raise it, and so I fail to raise my hand. Still, counterfactually, had I succeeded in raising my hand it would have been voluntary. *Mutatis mutandis* for rerunning belief processes. I can rerun a reasoning process hoping that it will result in a different belief, but even if it does not, that does not change my voluntarily undertaking the process of coming to believe.

Taking this into account, let us examine how S-BAR explains the existence of wishful thinking. On this account, wishful thinking is the output of the reasoning system, where the subject has tinkered with the process until they received the desired result and then stopped rerunning the process. Since outputs of the reasoning system are beliefs, wishful thinkings are beliefs. Subjects having all the evidence necessary to disbelieve a proposition need not disbelieve because they can re-evaluate the evidence until they get the desired response. While this may not happen consciously (remember that mode determination is determined by many internal and external factors), sometimes it does.

This discussion has implications for borderline cases of belief. For example, a mother who convinces herself into thinking that her son isn't doing drugs believes he's not doing drugs because her mental attitude is the finishing-point of a reasoning process. This belief may be somewhat fragile—if she starts to rerun the reasoning process it might be replaced with the more evidentially well supported position that her son does do drugs.

The voluntariness of belief can be elucidated by consider one issue that has come to light as a result of the replication crisis in the field of social psychology. A decade ago, some psychologists used of multiple statistical tools, opting for the statistical analysis that provided researchers with significance (see Simmons, Nelson, and Simonshohnn 2011, Francis 2012a, 2012b, John, Loewenstein, & Prelec 2012). Some psychology labs would run their data through one statistical model, and, if they did not get significance, would run it through another, and then

another, until they did get significance. The choice to rerun the reasoning problem is conscious (ask "given this data, is there a significant finding?"), as is mode determination, since the researchers actively chose the statistical model. Why rerun the statistics? Because the researchers wanted to find an effect. Plausibly, many psychologists believed that they had found an effect after the final rerun of the statistics. We know now that this is not good scientific practice because rerunning data on multiple models greatly increases the likelihood of false positives, but the use of multiple statistical analysis supports the voluntariness of some beliefs because at least some psychologists believed their results. Furthermore, S-BAR suggests that humans are like psychologists who rerun analyses because they can alter the mode of operation and control when they stop the reasoning process.[17]

Let's take a look at one representative argument against belief voluntarism from the nature of belief literature. Frankish (2007) argues against the psychological possibility of belief voluntarism because after a subject evaluates the evidence, if they want to then change their belief, they must change their belief about the nature of the evidence, but to do so requires changing a belief about the belief about the evidence. This chain continues *ad infinitum*, and since it is impossible for us finite creatures to complete an infinite series of cognitive acts it is impossible (psychologically) for us to believe voluntarily.

The crucial step in Frankish's argument is that the way one sees the evidence just is another belief. This is a common view, but if S-BAR is right, it is mistaken. There are our beliefs, acceptances, hypothesis, and other doxastic-like states, which are fed into the reasoning system. While our *evaluation* of the evidence—the product of the reasoning process—is a belief, that does not imply that the *evaluating* of the evidence is a belief or belief like state. Rather, the evaluating of a reasoning problem is not a representational state at all. Rather, it is a process. On S-BAR, that evaluating of the process consists in running and (possibly) rerunning the reasoning process. However, doing so does not 'add' more and more beliefs. A subject who reruns the reasoning process will generally not be adding higher-order beliefs about the evidence. Rather, they are revising—rewriting—already existing beliefs. Rerunning the process is more like a 'do-over' rather than adding amendments to a law. Since one is not adding beliefs in rerunning the reasoning process, Frankish's argument fails.

5.2 Belief and Acceptance

Beliefs and acceptance are generally distinguished by four properties: 1) acceptance is only held in a narrow domain compared to the wide domain of belief, 2) belief is involuntarily formed whereas acceptance is the result of a direct voluntary action, 3) acceptance is responsive to pragmatic, but not evidential, considerations, whereas belief is responsive to evidential rather than pragmatic considerations, and 4) in belief, but not acceptance, subjects are disposed to feel that *p* when considering *p*. On my view, the first property is the main way belief and acceptance are distinguished. On S-BAR, acceptances are cognitive quarantined, whereas beliefs are inferentially promiscuous. Relevant beliefs will typically be recruited (often automatically) for reasoning processes. This will not be the case for imaginings and acceptances. Because cognitive quarantining requires mental effort, there will be times when an imagining or acceptance 'leaks'

[17] Another analogy can be drawn between p-hacking and ordinary evidence-gathering, though nothing in the analogy is particular to S-BAR. P-hacking occurs when researchers continue to gather data while running statistical analysis and end the evidence gathering when significance is reached. Such a process is voluntary. Likewise, subjects can continue to gather evidence on some issue, ending the process when the evidence (subjectively) favors their desired result.

and so becomes a belief (or something very near to it). Because the account of kinds I adopt permits border cases and rejects necessary and sufficient conditions, this is not a problem for my taxonomy.

Why should we reject the other ways of distinguishing belief from acceptance? First, because S-BAR adopts a limited form of belief voluntarism, beliefs and acceptances cannot be distinguished etiologically by how they were formed. Second, the claim that belief is only responsive to evidential considerations, although perhaps right *normatively*, seems to fly in the face of the existence of motivated reasoning and wishful thinking. Because my method attempts to stick closely with actual (as opposed to ideal) human psychology, belief and acceptance shouldn't be distinguished primarily by normative considerations. This leaves the disposition for a feeling. I take 'feeling that p is true' in the case of belief to be a metacognitive attitude which psychologists call a 'feeling of rightness' (or FOR) (see Thompson 2009, Stanovich 2011). Others, such as Alston (1996) and Howard-Snyder (2013), also take the disposition for a FOR as a necessary condition for belief, and use this condition as a way to distinguish belief from acceptance (Alston 1996, p. 3-4). I have argued elsewhere (Mugg 2016b) that there are cases where beliefs lack this disposition—namely in cases of aversive racism and counterintuitive scientific and probabilistic claims (e.g. that the Copenhagen interpretation of quantum mechanics is true, or that it is more likely that an activist philosophy student became a bank-teller than a feminist bank-teller). On S-BAR, a worry about using FOR as a way of distinguishing belief from acceptance is that it is not clear that FOR is causally important on its own. If FOR is merely epiphenomenal, then it cannot be used to distinguish kinds, since, on the account of kinds I have adopted, kinds are distinguished causally. On the other hand, if beliefs possessing FOR are causally important because FOR implies that beliefs, more so than acceptances, will enter into reasoning processes when considered, then this causal property collapses into the property that belief, rather than acceptance, is inferentially promiscuous as I have suggested.

5.3 Indirect Measurement

Inferential promiscuity of belief implies that belief is tied to automaticity in a way that acceptance is not. If this is right, then it is possible to measure some beliefs using indirect measurements. Of course, the inverse is not true: not all indirect measurements in psychology measure beliefs (e.g. the IAT may not measure beliefs). Interestingly, S-BAR offers a disambiguation between two distinct indirect methods in psychology: time constraint and cognitive load. In the case of time restraint, what psychologists would be measuring is an initial (or very early) output of the reasoning system. Depending on the mode of operation, the subject would plausibly only recruit the most seemingly relevant beliefs to the reasoning process. However, in this case it is also possible that we are not capturing the previously held beliefs of the subject, and are instead only capturing a truncated reasoning process. Since belief is the output of a reasoning process, this may actually induce a belief rather than uncover an existing one. In particular, this would be the case if the subject stopped reasoning after they had delivered their time-constrained response. In the case of cognitive load, we are not allowing the reasoning system to enter into modes of operation that require higher cognitive effort, such as novel thinking. Supposing that subject did already have beliefs formed about the content of which they are asked, their inability to enter into modes of operation requiring more working memory will require them to respond with beliefs already held rather than form new ones. Thus, psychologists wishing to uncover existing beliefs should use cognitive load methods over response time.

6. Conclusion

We can think of belief as a cognitive kind because it is a node in a causal network. Here I've outlined one cognitive architecture, examined how belief should be understood on this picture of human reasoning, and suggested some implications for the nature of belief and indirect measurement in psychology. Of course, much work remains for fleshing out the doxastic taxonomy on S-BAR, examining the nature of belief in non-reasoning processes (such as perception), and, perhaps most importantly, evaluating the empirical adequacy of the cognitive architecture I have offered here.

Acknowledgements: Thank you to Devin Curry, Tim Crane, and the editors of this volume for helpful feedback on drafts of this paper. Thank you to participants at the University of Helsinki's 2023 Dynamics of Agency Workshop for feedback on section 3.

Work Cited

Alston, W. (1996). 'Belief, Acceptance, and Religious Faith,' in J. Jordan and D. Howard-Snyder (Eds.), *Faith, Freedom, and Rationality.* Lanham, MD: Lowman & Littlefield, 3–27.

Boyd, R. (1989). "What Realism Implies and What it Does Not," *Dialectica* 43(1–2):5–29.

Boyd, R. (1999). "Homeostasis, Species, and Higher Taxa," in R. A. Wilson (Ed.), *Species: New Interdisciplinary Essays*. Cambridge, MA: MIT Press,141-85.

Carruthers, P. (2009) "An Architecture for Dual Reasoning," in Evans, J. St. B. T. and Frankish, K. (Eds.), *In Two Minds: Dual Processes and Beyond*. Oxford University Press, 109–127.

Cohen, L. J. (1992). *An Essay on Belief and Acceptance*. New York: Clarendon Press.

Colaço, D. (2022). "What Counts as a Memory? Definitions, Hypotheses, and 'Kinding in Progress,' " *Philosophy of Science* 89(1):89-106.

Curry, D. S. (2021a). "*g* as Bridge Model," *Philosophy of Science* 88(5):1067-1078.

Curry, D. S. (2021b). "How Beliefs are Like Colors," *Synthese* 199(3-4):7889-7918.

Craver, C. F. (2009). "Mechanisms and Natural Kinds," *Philosophical Psychology* 22(5):575-594.

Dale, M. T. (2020). "Neurons and Normativity: A Critique of Greene's Notion of Unfamiliarity," *Philosophical Psychology* 33(8):1072-1095.

Evans, J. S. B. (2011). "Dual-Process Theories of Reasoning: Contemporary Issues and Developmental Applications," *Developmental Review 31*(2): 86–102.

Evans, J. S. B. and K. E. Stanovich. (2013). "Theory and Metatheory in the Study of Dual Processing Reply to Comments," *Perspectives on Psychological Science 8*(3): 263–271.

Evans, J. S. B., J. Barston, and P. Pollard. (1983). "On the Conflict Between Logic and Belief in Syllogistic Reasoning," *Memory & Cognition* (11):295-306.

Francis, G. (2012a). "The Psychology of Replication and Replication in Psychology," *Perspectives on Psychological Science* 7(6):580-589.

Francis, G. (2012b). "Publication Bias and the Failure of Replication in Experimental Psychology," *Psychonomic Bulletin & Review* 19(6):975-991.

Frankish, K. (2004). *Mind and Supermind*. Cambridge: Cambridge University Press.

Frankish, K. (2007). "Deciding to Believe Again," *Mind* 116(463):523-547.

Gawronski, B. and J. De Houwer. (2014). "Implicit Measures in Social and Personality Psychology," in H. T. Reis, and C. M. Judd (Eds.), *Handbook of Research Methods in Social and Personality Psychology* (2nd edition). New York, NY: Cambridge University Press.

Gawronski, B. and E. P. LeBel. (2008). "Understanding Patterns of Attitude Change: When Implicit Measures Show Change, but Explicit Measures do Not," *Journal of Experimental Social Psychology* (44):1355-1361.

Gendler, T. S. (2008a). "Alief and Belief," *Journal of Philosophy* 105(10):634-663.

Gendler, T. S. (2008b). "Alief in Action (and Reaction)," *Mind and Language* 23(5):552-585.

Gigerenzer, G. (2018). "The Bias Bias in Behavioral Economics," *Review of Behavioral Economics* (5)3-4: 303-336.

Gigerenzer, G. (2010). "Personal Reflections on Theory and Psychology." *Theory & Psychology*, *20*(6),733–743.

Gilbert, D. (1991). "How Mental Systems Believe," *American Psychologist* 46(2):107–19.

Gilovich, T. and D. Griffin. (2002). "Heuristics and Biases: Then and Now," in T. Gilovich, D. Griffin, and D. Kahneman (Eds.), *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge: Cambridge University Press.

Gomez-Lavin, J. (2021). "Working Memory is Not a Natural Kind and Cannot Explain Central Cognition," *Review of Philosophy and Psychology* 12(2):199-225.

Greene, J. (2014). "Beyond Point-and-Shoot Morality: Why Cognitive (Neuro)science Matters for Ethics," *Ethics* 124(4):695–726.

Greenwald, A. and M. Banaji. (1995). "Implicit Social Cognition: Attitudes, Self-esteem, and Stereotypes," *Psychological Review* 102(1):4-27.

Hieronymi, P. (2006). "Controlling Attitudes," *Pacific Philosophical Quarterly* 87(1):45-74.

Hieronymi, P. (2008). "Responsibility for Believing," *Synthese* 161(3):357-373.

Howard-Snyder, D. (2013). "Propositional Faith: What it is and What it is Not," *American Philosophical Quarterly* 50(4): 357–372.

Howard-Snyder, D. (2016). "Does Faith Entail Belief?," *Faith & Philosophy* 33(2):142–162.

Howard-Snyder, D. (2019b). "Three Arguments to Think that Faith Does Not Entail Belief," *Pacific Philosophical Quarterly* 100(1):114–128.

Inbar, Y. and J. Lammers. (2012). "Political Diversity in Social and Personality Psychology," *Perspectives on Psychological Science* 7:496-503.

Jennings, C. D. (2020). *The Attending Mind*. New York: Cambridge University Press.

John, L. K., G. Loewenstein, and D. Prelec. (2012). "Measuring the Prevalence of Questionable Research Practices With Incentives for Truth Telling," *Psychological Science* 23(5):524–532.

Kahneman, D. (2011). *Thinking, Fast and Slow*. New York: New York: Farrar, Straus and Giroux.

Kornblith, H. (1993). *Inductive Inference and its Natural Ground: An Essay in Naturalistic Epistemology*. Cambridge, MA: MIT Press.

Khalidi, M. A. (2017). “Crosscutting Psycho-neural Taxonomies: The Case of Episodic Memory,” *Philosophical Explorations* 20(2):191-208.

Khalidi, M. A. (2018). “Natural Kinds as Nodes in Causal Networks,” *Synthese* 195(4):1379-1396.

Khalidi, M. A. (2023). *Cognitive Ontology: Taxonomic Practices in the Mind-Brain Sciences.* Cambridge: Cambridge University Press.

Khalidi, M. A. and J. Mugg. (2023). “Cognitive Heuristics and Biases,” in *Cognitive Ontology: Taxonomic Practices in the Mind-Brain Sciences.* Cambridge: Cambrige University Press

Keren, G. and Y. Schul. (2009). "Two is Not Always Better Than One: A Critical Evaluation of Two-System Theories," *Perspectives on Psychological Science 4*(6):533–550.

Keren, G. (2013). "Tale of Two Systems: A Scientific Advance or a Theoretical Stone Soup? Commentary on Evans & Stanovich (2013)," *Perspectives on Psychological Science* 8(3):257–262.

Kruglanski, A. W. (2013). "Only One? The Default Interventionist Perspective as a Unimodel—Commentary on Evans & Stanovich (2013)," *Perspectives on Psychological Science 8*(3):242–247.

Kruglanski, A. W. and G. Gigerenzer. (2011). "Intuitive and Deliberate Judgments are Based on Common Principles," *Psychological Review 118*(1):97.

Levy, N. and E. Mandelbaum. (2014). "The Powers that Bind: Doxastic Voluntarism and Epistemic Obligation," in J. Matheson (Ed.), *The Ethics of Belief*. New York: Oxford University Press, 12-33.

Leslie, A. M. (1987). "Pretense and Representation: The Origins of 'Theory of Mind'," *Psychological Review* 94(4):412-426.

Malcolm, F. and M. Scott. (2017). "Faith, Belief, and Fictionalism," *Pacific Philosophical Quarterly* 98(S1):257–274.

Mandelbaum, E. (2013). "Against Alief," *Philosophical Studies* 165 (1):197-211.

Mandelbaum, E. (2014). "Thinking is Believing," *Inquiry* 57(1):55-96.

McKaughan, D. (2013). "Authentic Faith and Acknowledged Risk: Dissolving the Problem of Faith and Reason," *Religious Studies* 49(1):101-124.

Monsell, S. and J. Driver. (2000). "Banishing the Control Homunculus," *Control of Cognitive Processes: Attention and Performance XVIII*. Cambridge, MA: MIT Press, 3-32.

Mugg, J. (2016a). "The Dual-Process Turn: Why Recent Defenses of Dual-Process Theory of Reasoning Fail," *Philosophical Psychology* 29(2):300-309.

Mugg, J. (2016b). "A Defense of the Belief-Plus Model of Faith," *European Journal of Philosophy of Religion 8(2):201-219.*

Mugg, J. (2018). "The Sound-Board Account of Reasoning: A One-System Alternative to Dual-Process Theory," *Philosophical Psychology* 31(7):1046-1073

Mugg, J. (2022). "Faith Entails Belief: Three Avenues of Defense Against the Argument from Doubt," *Pacific Philosophical Quarterly* 103(4):816-836.

Nichols, S. and S. P. Stich. (2003). *Mindreading: An Integrated Account of Pretense, Self-awareness, and Understanding Other Minds.* New York: Clarendon Press/Oxford University Press.

Nosek, B. A., C. B. Hawkins, and R. S. Frazier. (2011). "Implicit Social Cognition: From Measures to Mechanisms," *Trends in Cognitive Sciences* 15:l52-159.

Olson, M. A. and R. H. Fazio. (2006). "Reducing Automatically Activated Racial Prejudice Through Implicit Evaluative Conditioning," *Personality and Social Psychology Bulletin* 32:421-433.

Porot, N. and E. Mandelbaum. (2020). "The Science of Belief: A Progress Report," *Wiley Interdisciplinary Reviews. Cognitive Science*, e1539:1-17.

Rollwage, M., L. Zmigrod, L. de-Wit, R. J. Dolan, and S. M. Fleming. (2019). "What Underlies Political Polarization? A Manifesto for Computational Political Psychology," *Trends in Cognitive Sciences 23*(10):820-822.

Samuels, R. (2009). "The Magical Number Two, Plus or Minus: Dual-Process Theory as a Theory of Cognitive Kinds," in Evans, J. St. B. T. and Frankish, K. (Eds.), *In Two Minds: Dual Processes and Beyond*. Oxford University Press, 129–146.

Sloman, S. A. (1996). "The Empirical Case for Two Systems of Reasoning," *Psychological Bulletin 119*(1):3–32.

Shah, N. and D. Velleman, J. (2005). "Doxastic Deliberation," *Philosophical Review* 114 (4):497-534.

Stanovich, K. (2011). *Rationality and the Reflective Mind*. Oxford: Oxford University Press.

Stanovich, K. E., M. E. Toplak, and R. F. West. (2020). “Intelligence and Rationality,” in R. J. Sternberg (Ed.), *The Cambridge Handbook of Intelligence*. Cambridge: Cambridge University Press, 1106–1139.

Stanovich, K. E., R. F. West, and M. E. Toplak. (2016). *The Rationality Quotient: Toward a Test of Rational Thinking.* Boston: Boston Review.

Thompson, V. (2009). “Dual Process Theories: A Metacognitive Perspective,” in Evans, J. St. B. T. and Frankish, K. (Eds.), *In Two Minds: Dual Processes and Beyond*. Oxford University Press, 171–196.

Wimsatt, W. (1994). “Ontology of Complex Systems: Levels of Organization, Perspectives, and Causal Thickets,” *Canadian Journal of Philosophy* 20(sup1):207-274

Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

Wu, W. (2016). “Experts and Deviants: The Story of Agentive Control,” *Philosophy and Phenomenological Research* 93(1):101-26.