Exploring by Believing

Word Count: 15,233

May 25, 2020

Abstract

Sometimes, we face choices between actions most likely to lead to valuable outcomes, and actions which put us in a better position to learn. These choices exemplify what is called the *exploration/exploitation trade-off*. In computer science and psychology, this trade-off has fruitfully been applied to modulating the way agents or systems make choices over time. In this paper, I argue that the trade-off also extends to belief. We can be torn between two ways of believing, one of which is expected to be more accurate in light of current evidence, whereas the other is expected to lead to more learning opportunity and accuracy in the long run. Further, it is sometimes rationally permissible to choose the latter. I break down the features of action which give rise to the trade-off, and then argue that each feature applies equally well to belief. This conclusion is an instance of a systematic, foreseeable way in which what is rational to believe now depends on what one expects to be doing in the future. That is, epistemic rationality fundamentally concerns *time*.

Prelude

The physiologist Ivan Pavlov spent most of his career split between two ways of living: during the academic year, he would be hard at work in his laboratory, and in the summers, he put aside all of his scientific research and read fiction at his dacha (summer house). We could imagine that these two ways of living came along with two ways of thinking. During the academic years, he might have really believed that he could understand the mind through studying nerves (an idea he called "nervism"). During the summers, reading literature with spiritual themes, he might have instead believed that the science of the soul faced insurmountable obstacles. While it seems possible that switching one's habits based on the rhythm of the academic year is a rational way to live, switching back and forth between two contradictory ways of believing based on the time of year seems irrational. Could there be a good reason for these seasonal shifts of belief? To answer this question, we'll first need to understand why seemingly arbitrary

switches between two ways of *living* might be rational, and how switches in living relate to and differ from switches in belief.

1 Introduction

In many decision-making scenarios, we can observe a trade-off between choosing the action that maximizes expected reward, or the action most likely to result in learning something new: the **exploration/exploitation trade-off**. For instance, imagine you are choosing between ordering your favorite ice cream flavor or trying a new one. Exploiting consists in picking the option most likely, on your evidence, to have the highest value. Exploring, on the other hand, involves choosing something previously untested or about which you're uncertain. There's a trade-off because the best behavior for exploring (say, trying every flavor once, even banana-tomato) is rarely the behavior that is the most likely to maximize reward - and vice versa. The striking result, in the case of action, is that these exploratory behaviors that look like seeking out costly information are rationalized entirely without appealing to an agent who values information itself; even if I only love tastiness, I should still sometimes try flavors that seem likely to be disgusting. The task of this paper is to extend the idea of such a trade-off to the case of belief formation and change: should we ever believe solely in order to explore?

Initially, the prospect of a symmetry between exploration in action and exploration in belief might look unlikely. For one, actions are chosen voluntarily, whereas beliefs are arguably formed without an act of the will (see [5] for a discussion of this question). So the exploration/exploitation trade-off might be a decision-theoretic concept that is out of place in the context of belief. Likewise, we usually think of epistemic rationality as universal and unchanging, whereas rational decision-making allows for trade-offs and merely instrumental rational actions.

However, I will argue that there is indeed an exploration/exploitation trade-off in belief, because of the connection between our current beliefs and our dispositions to conduct experiments and explore the space of possibilities. This paper is the first to posit exploration in belief. However, others have argued for exploration in other parts of cognition, for instance Sripada [44] argues for exploration in the act of imagination. While past work on epistemic trade-offs in belief has focused on situationally-driven trade-offs that are arbitrary and often

fantastical [19], this paper looks at a learning situation that characterizes a large portion of our epistemic position in real life, and posits a systematic and easily implemented rule: deviate occasionally from the recommendations of your best belief change policy in the beginning of inquiry.

The beginning of inquiry is determined by how long the inquiry will extend into the future as compared to how long it has progressed so far, and how much more evidence will be acquired as compared to how much evidence has already been collected, among other things. These features are based in more than just evidence; two agents may have exactly the same evidence about some issue, but if one expects to get more evidence in the future than the other, they may be in different stages of inquiry. This is significant because epistemologists have long assumed (whether implicitly or explicitly) that considerations about what the agent will be doing in the future, and how long they'll have to do it, are irrelevant to epistemic rationality. Consider, for instance, how unlikely it is for a typical case in the literature on peer disagreement to mention what further evidence might be available after the current episode. Along similar lines, convergence arguments (like the one debated in [8] and [21], or [26]) ground rational procedures by appealing to the limiting case of infinite evidence. Consequently, one of the goals of this paper is to propose that we attend more seriously to facts about the agent's evidential position over time.

The structure of the paper is as follows: in § 2, I survey the formal literature on the exploration/exploitation trade-off in action and identify some structural features of decision problems which give rise to the trade-off. In § 3, I introduce an example of belief which I'll use to demonstrate what exploration in belief might look like. I then delineate the analogy more precisely in § 4. The core argument, in § 5, appeals to how belief rationally guides and constrains imagination. § 7 discusses objections and the significance of this project and analyzes its relation to other questions in epistemology including epistemic consequentialism.

2 The Exploration/Exploitation Trade-Off

The exploration/exploitation trade-off is a formal device grounded in a constrained and simplified decision problem. The disadvantages of using such a device to make a point about human rationality are serious: to find an approximation of a real agent and a real decision problem that

we can apply the trade-off to risks idealizing away from important pieces of the actual problem. Even given all this, I employ this approach for two reasons.

First, existing work on epistemic trade-offs has focused exclusively on one-off situations, where a trade-off means making a choice at a particular time-point between a set of well-delineated options. On the contrary, a formal approach can describe a much more general kind of trade-off. Exploration/exploitation, and other formal trade-offs such as bias/variance or depth/precision, describe dynamics that apply to all members of a broad class of approaches, as well as to iterations over time. In the philosophical context, this generality is significant; as I'll discuss later, a general trade-off, unlike a specific one, faces fewer worries about feasibility under uncertainty, it can be derived from a creature-construction perspective, and has far more direct implications for cognitive science and psychology. This advantage of generality has been utilized, for instance, in recent work on trends in scientific communities by Thoma [46], O'Connor & Bruner [37], and Mayo Wilson [33] among others.

Second, idealization, when used correctly, allows us to see the bare bones of a given situation. The aim of this paper is to illuminate a possible similarity between the rationality of belief and that of action. As such, viewing both action and belief through the lens of the same formal problem will bring into view both similarities and differences between the two, as we are able to see both where the formal framework is a good fit, and where it fails to capture features of interest. Both of these advantages of formal modeling rely crucially on understanding both when a model is applicable, as well as when it misses something crucial – we can often learn from attending to either, as I will attempt to do in what follows.

In § 2.1, I explain the trade-off through a classic setup in the literature: the multi-armed bandit. Readers familiar with the trade-off may skip to § 2.2, in which I present an original derivation of the conditions under which the trade-off applies.

2.1 The Multi-Armed Bandit

In this example, I review results that show that *as a rule*, in order to receive the optimal reward from many environments, a (somewhat limited) agent should occasionally choose actions not recommended by her best policy. By as a rule, I mean that this result predictably applies to many environments, and that one could reasonably believe that choosing a non-recommended

action would help based on limited information. Optimal reward will be measured by aggregate preference satisfaction, which in this toy example will be total number of dollars won. This section serves two functions: (a) it explicates the exploration/exploitation trade-off for action, and (b) it establishes that some behaviors that seem to reflect a preference for information are rationalizable for agents who do not intrinsically value information. That is, if exploring non-recommended options is predictably associated with optimal reward, then rational agents will carry out these behaviors regardless of what they take to be optimal reward.

We'll start with a simple expected utility (EU) framework. We have some agent, who has a probability distribution representing her credences over various outcomes and combines these with a corresponding utility function to generate expected utilities. Canonically, these outcomes are complete states of the world. However, in practice, we often idealize away from these complex states into simpler ones, and evaluate only the value of the immediate result of each action. For now, our expected utility framework will be near-sighted or *myopic* in this way.

Now here's the problem our agent faces. She can choose to play at one of three slot machine arms $i-k$. After each play, she may continue at the same arm, or switch to a different arm. Each arm produces stochastic rewards distributed around a fixed unknown bias [1]. Let's say she starts with the following estimations of the biases, where a higher bias means a higher probability of a valuable outcome: $b_i = .5, b_j = .2, b_k = .1$. Now, assuming that she's going to play these slot machines for some significant amount of time, what should she do?

One method would be to always choose the arm with highest expected reward, calculated from the estimated bias and her confidence in that estimate. She would start by choosing arm i. After she plays i, she'll get some information. Let's say that the true bias of i is .8, and the outcomes in the short term reflect that bias fairly faithfully. So by using this method, she will continue to choose i over and over again, because its estimated payout will never drop below that of arm j, which has the next highest estimate. This is the method recommended by her myopic expected utility rule: the myopic exploitation policy. It's exploitative because it always does what is best according to current expectations.

How good is the myopic exploitation policy? If she's right initially that i is the best arm,

[1] Multi-armed bandit problems tend to have looser assumptions around bias, for instance that the reward state evolves according to some unknown Markovian function [31]

she'll attain the optimal reward. However, if she's wrong, and for instance k actually has a bias of .9, her total reward will not be optimal, and indeed will be significantly suboptimal as the choice is repeated over and over. She has no reason to try the other arms if she only acts to maximize reward at the next step. Myopic exploitation has a significant risk of getting her stuck in a local maximum, a section of the reward landscape that is better[2] than all neighboring possibilities but not the best overall. Once she's in the bandit situation described above, she'll never stop making the same suboptimal choice.

A very simple way of allowing for exploration in an exploitative decision strategy (where A is the act with highest exploitative value) is to add this rule: at every decision point, choose a random act other than A with probability ϵ, or choose A with probability $1 - \epsilon$. This is called an ϵ-greedy strategy. As we increase ϵ, our agent explores more. As ϵ approaches 1, our agent will learn a lot, but her learning will not benefit her, since her knowledge about the options won't affect her behavior at all. As ϵ approaches 0, her behavior will converge to the myopic exploitation policy. Because she will learn more and more about her environment as she makes these choices, it's reasonable for her to start off exploring a lot and then exploit more and more as information accumulates - when she knows everything about the outcomes, there's no need to try new things, whereas when her expectations are poorly informed, maximizing expected utility is unlikely to be particularly effective.

As it happens, ϵ-greedy methods approximate optimal solutions to many bandit-style problems. In some problems, a fully optimal solution can be arrived at by calculating the Gittins index of each arm, which is roughly the value of continuing to use that arm adjusting for the potential of learning. This approach splits the high-dimensional optimization problem into a series of more tractable problems: the Gittins index for an arm is re-calculated only when that arm is chosen. [3] Thus the Gittins index policy is optimal when the bias of each arm is inde-

[2] 'better' would be filled in with whichever candidate for epistemic value we ultimately decide on.

[3] Solving the problem involves calculating the index for each arm i , given by the following equation:

$$v^i(x^i) = \max_{t>0} \frac{\mathbb{E}[\sum_{t=0}^{\tau} \beta^t r^i(X_t^i) | X_0^i = x^i]}{\mathbb{E}[\sum_{t=0}^{\tau} \beta^t | X_0^i = x^i]}$$

Where τ is a stopping time, r is a reward, $x^i \in X^i$ is a state and β is the survival probability, which is the probability that the situation continues into the next iteration. Then, the optimal policy is to always play the arm with the highest Gittins index. This is a computationally expensive procedure (relative to approximations such as Upper Confidence Bound [6] (UCB) and ϵ-greedy Q-learning) that relies on forward induction [31]. Crucially, the Gittins index of each arm typically declines after each play, so the agent does not continue to play the same arm

pendent from all other arms and does not change over time. These, of course, are substantial assumptions that may not hold in more realistic situations.

2.2 From Bandits to Everyday Choice

Having laid out the basics of this trade-off in a contrived, formal context, I'll now discuss how this piece of machinery applies to practical rationality more generally. This requires understanding what kind of agents are well-described by the trade-off, and in what kinds of decision problems it applies.

The first issue concerns the properties of the agent: does the rationality of exploration only hold because of failures particular to myopic expectations? Within the category of approximations of fully ideal agents, we can observe examples of the rationality of exploration, and the applicability of the trade-off more generally. One series of examples comes from reinforcement learning (RL) [45], a family of learning algorithms that have been used to model animal and human decision-making, among other applications. In RL, the agent calculates the value of each progressive step that she might possibly take, multiplied by a discounting factor [4]. Reinforcement learning algorithms can plan over arbitrarily many future steps, and yet standard models include perturbations designed to induce exploration. Exploration is explicitly encoded in a wide range of RL methods, from basic algorithms such as Q-learning to more complex model-based methods. So the exploration/exploitation trade-off does not require myopia.

Moving to a truly ideal agent, we might ask about the trade-off in an orthodox decision theory context where agents are capable of planning to an infinite horizon. One option might be to treat the trade-off as a heuristic for describing the behavior of such agents, even though these agents always maximize expected value and so do not in a deeper sense trade anything off (see [22] for an in-depth treatment of relaxing idealization and its consequences for stocasticity in choice, among other things). [5] Whatever we might want to say about such a fully ideal agent, however, will not carry over into the epistemic case, since the learning problem I will discuss

even if it generates high reward.

[4] This sometimes includes every possible act, or is cut off at a future horizon - see [25] for arguments that employing a horizon may actually be optimal

[5] On the other hand, Rothschild [41] proves that there is a positive probability that an EU-maximizing agent will settle on a policy of choosing the wrong arm and continue that way forever. Whether this proof implies that ideal agents should sometimes diverge from EU-maximizing behavior is unclear. First, the proof makes a few crucial assumptions about the agents. Second, a proponent of classical EU might argue that this is merely a case of subjective rationality that is unfortunately punished by an unlearnable environment.

is one of discovering new hypotheses. Because of this, our agent will necessarily be bounded.

Second, what situations give rise to this trade-off? The idea of exploration versus exploitation has been used in various fields in a wide variety of situations: for instance in clinical trials [38], developmental psychology [18], neuroscience [52], and to describe foraging behavior in birds [30]. In deciding how to collect food in some landscapes, an animal benefits from deviating from the strategy of choosing the patch that looks likely to contain the most food. In other landscapes, there is no reason to explore, and the animal should always exploit. And in still others, there is nothing that could be called exploring at all. So how can we tell the first kind of environment from the others?

There are two ways to argue that the exploration/exploitation trade-off applies in a new case without modeling competing methods directly. First, we might demonstrate that the new case is merely a superficial transformation of an old case. For instance, we might re-describe the case of clinical trials as a multi-armed bandit problem. However, until we can demonstrate better outcomes through the use of an exploratory strategy in the new case, it is in principle possible that any re-description elides relevant differences between the cases. A second strategy is to derive general features that seem to apply to most or all cases in which the trade-off obtains. If these features obtain in the new case, then we have a reason to expect the trade-off to obtain in the new case as well. That is, the first strategy relies on a one-to-one similarity between cases, the second on categorizing the new case according to features observed across a wide range of past cases.

In pursuit of the second strategy, I'll provide a brief gloss on four features of decision problems that lead to a meaningful application of the trade-off. Considering the foraging environments, any setup with only a handful of chances to collect food will not be solved by exploration, since there won't be enough future chances to put new information to work. The animal must not already have enough information to understand the relevant features of the environment or exploration will not be beneficial; conversely, they must have some information about the environment, or exploitation will not be meaningful. The problem must involve uncertainty, but not be totally blind. The animal's behavior must be systematically linked with acquiring evidence; that is, there must be behaviors that predictably raise the probability of getting new evidence, otherwise exploration would be impossible. More interestingly, these behaviors must

not line up perfectly with the behaviors that generate value. It can't be that feeding from each patch is always (predictably) good for getting new evidence in proportion to how (predictably) good it is for getting more food. The problem must be a sequential one, with a sufficient number of iterations. In short, the exploration/exploitation trade-off is meaningful when reward and evidence are both linked to acts (conditions 1 & 2), the decision problem is iterated over and over (condition 3), and the degree to which an act generates reward diverges from the degree to which the same act generates evidence (condition 4). I will return to these conditions in § 4.

A third issue is to isolate key features of the trade-off that will be crucial for belief. The critical feature will be the relationship between exploration/exploitation and time. As I noted at the end of § 2.1, there's a somewhat generic rationale for preferring to explore more earlier and exploit more later. This reflects a relationship between time and uncertainty, since exploration is more important when uncertainty is high. However, even while holding uncertainty fixed, we find a relationship between exploration and time. The information which is reached by exploring has more value when our agent will have a lot more chances to play the slot machines. As she approaches the end of her interaction with the current environment, the diminishing of future opportunities favors exploitation. This is so even if she is still quite uncertain. Take two agents who are equally uncertain, one pulling the first lever of a long sequence and the other pulling the final lever. The first agent has more reason to explore than the second. Of course, in a real-life situation, the boundaries of one context are not given objectively from the world but the agent herself plays a role in defining what counts as the same problem, and in acting in ways that change how problems extend over time.

Consider this lyric from a Frankie Ballard song: 'how am I ever gonna get to be old and wise, if I ain't ever young and crazy?". This expresses a common sense version of the idea behind the trade-off. When you're young, you have an extra reason to act crazy - or to deviate from the action that looks like the best bet from a strategic perspective. The best action for learning is not always the most subjectively rational. Here, the modulation of the trade-off over time is front and center. Young Frankie Ballard should say yes to things that old Frankie Ballard should not, for the same reasons as in the formal case.

Reward changes in the environment also modulate the trade-off. Traca and Rudin [48] show that in environments with periodic reward functions, it's optimal to exploit more during high-

reward periods and explore more during low-reward periods. In their case, the varying rewards were due to daily traffic patterns on a website, and at higher traffic times, the recommender algorithm did best by exploiting more, and by exploring more at lower traffic times. In summary, variations in uncertainty, potential for actions, and total available reward all modulate the exploration/exploitation trade-off in action.

3 A Case of Exploratory Belief

3.1 Why Belief?

I'll now turn to the case of belief[6]. This involves changing our focus from practical value, e.g. attaining dollars from a slot machine, to epistemic value, e.g. acquiring an accurate model of an environment. That is, while it might make sense to value whichever beliefs will make me the most money, I'm interested in the kind of value beliefs have when they are true, regardless of their usefulness.

An exploration-exploitation trade-off in belief would have three significant consequences. First, just like exploring by choosing a flavor at random could be rational, believing by adopting a belief at random might be rational. Second, believing with the greatest accuracy *now* would sometimes be at odds with getting to the most accurate belief in the long run. Third, the rationality of belief would depend on where the agent is in the process of inquiry, even holding fixed all of her direct evidence about the situation at hand – just as two players at the slot machine who've just observed the exact same sequence of pay-offs should choose differently depending on whether the current play is their last or one of many future plays. All of this would hold without adding in any new source of epistemic value beyond accuracy or truth.

All of this suggests that an exploration/exploitation trade-off in belief would shake up the classic debate in epistemology between James [23] and Clifford [12] over the reasons to believe. Roughly, Clifford's position was that belief should always be based on the evidence, and undertaken for pure epistemic reasons, whereas James held that we should sometimes believe beyond what the evidence can support for practical reasons. The exploration/exploitation trade-off for

[6] A different version of this paper would target credences instead of full belief. This would have the advantage of more precision, but full belief is accepted as the subject of epistemology by a larger set of scholars. While I can't go into details here, the argument for credences would not target probabilistic incoherence but instead incoherence in the representations themselves.

belief would mean there is a third position available. Sometimes, purely epistemic reasons that have nothing to do with our values, desires or interests support beliefs at odds with the current evidence

3.2 An Example

Our question is this: do agents who do well at acquiring epistemically valuable beliefs exhibit the exploration/exploitation trade-off? To do so, let's return to the initial story about Ivan Pavlov. To be clear that this is not meant to be an analysis of the historical Pavlov's actual psyche, I'll refer to our fictionalized Pavlov as Vanya. Vanya is facing an epistemic challenge: he is in deep uncertainty about the nature of the connection between the mind and the body. He's considering two hypothesis which seem to him to be mutually exclusive. The first, nervism, dictates that the science of nerves will ultimately be able to explain thought. The second, mysticism, holds that there is an ineffable element to human thought such that we can conceive of minds without bodies, and because of this, thought must be studied through first-personal reflection. I'll stipulate that Vanya's evidence about this question is such that it supports suspension of judgment[7] ; he does not know enough to decisively conclude one or the other option is true. However, instead of adopting this evidentially supported response, Vanya switches between believing these two hypotheses[8] .

> **Conflicted Vanya**: Vanya is receiving conflicting evidence about the nature of the mind. He responds by believing that the human mind is entirely a product of material changes to nerves during the academic year, and believing that the mind is essentially outside the natural material order when he spends his summers in the dacha. These switches are not brought on by changes in evidence.

Does this mean that Vanya could have one coherent credence function underlying this seemingly incoherent behavior? Having a partial (but coherent) credence in the propositions in question might sometimes express itself in belief-like and disbelief-like behaviors depending on the stakes. However, we would expect these behavioral switches to correspond to switches in the

[7]Suspension of judgment is a doxastic attitude distinct from belief and disbelief characterized by the agent stopping short of coming to a verdict about the truth of some proposition (the status of suspension as an attitude is somewhat controversial, see [17] for discussion).

[8]I assume here and throughout that belief and suspension of judgment are mutually exclusive attitudes. If one thought that a weak form of belief was compatible with suspension of judgment (perhaps motivated by considerations like those raised by Hawthorne et al [20]), then the argument I give here should be taken to contrast suspension of judgment with high credence or strong belief.

stakes and/or evidence, and by stipulation, Vanya is oscillating back and forth by following an internal routine rather than an external shift[9] .

Further, Vanya will not be in this divided state forever; instead, he's developing two incoherent projects in parallel in order to eventually be able to figure out which is better. Since the belief in nervism or mysticism is a foundation linked to many other beliefs, and which often determines how other beliefs are evaluated, it's reasonable to think that either future coherent state will have very different standards of evaluation and recommend distinct experiments. So Vanya is also in a state of meta-uncertainty or uncertainty about the right standards to apply to his beliefs and evidence, which he responds to by moving to a less warranted state (in this case by *all* standards) that might improve his prospect of learning.

In what follows, I'll discuss the reasons Vanya might have to adopt the switching policy versus the suspension of belief policy, and connect this case to the exploration/exploitation trade-off. Note that both of these policies directly govern only the answer to a particular question, and can be thought of as concerning a single belief: that nervism is true and materialism false. This belief is what we might call a framework belief: a belief that grounds, or can be expected to ground, a large set of other beliefs.

4 Belief/Action Symmetry

In this section, I argue that the multi-armed bandit and the problem Vanya faces contain essentially the same structure. Therefore, the exploration/exploitation trade-off is operative in both cases. More generally, I present a view on which the trade-off should be a normal feature of good reasoning about what to believe. In doing so, I'll highlight both similarities and differences between action and belief, and between the highly constrained bandit problem and more realistic cases of belief and action.

To make this argument, I'll sketch the features of the bandit case, and then extend them to the case of belief. It's important to note here that parity between the bandit and the context of belief is a stronger criterion than necessary to establish the existence of an exploration/exploitation trade-off in belief; multi-armed bandits are one of many problems whose solutions exhibit this trade-off ranging from tree search to applied problems in robotics.

[9] See [10] and [47] for more discussion of diachronic norms that prohibit non-evidence-driven changes in belief or credence.

Here's an overview of these background conditions:

	action bandit	*belief bandit*
Condition 1: generates evidence:	usually	usually
Condition 2: generates reward:	sometimes	sometimes
Condition 3: procedure is iterated:	approximately	approximately

Classic bandit problems exhibit a trade-off because we expect pulling the lever to give us evidence about the underlying function, and also a reward. The process needs to be iterated - otherwise exploitation would always trump exploration.

The three background conditions make the following critical condition possible:

	action bandit	*belief bandit*
Condition 4: evidence and reward diverge:	sometimes, progressively	sometimes, progressively

For instance, in the case of a high-reward lever that has been pulled many times, so that one more pull will likely provide little evidence but a lot of money. I aim to show that all of these features can be found in our everyday problem of what to believe: belief changes what kinds of evidence we can expect to receive based on our dispositions to imagine and conduct experiments, it gives us a 'reward' in the form of accurate beliefs, and these two diverge in cases like Vanya's that involve framework beliefs. Each act of forming a belief, like each action, is in some sense perfectly unique. But in both cases, we're engaged in a complex process that can be approximated for some purposes by treating it as a series of iterated moves.

A background issue here concerns internalism and externalism. An externalist version of these conditions would require that, for instance, reward and evidence actually generally come apart, regardless of whether the agent could plausibly be expected to know that fact. On the other hand, a standard internalist version would require that the agent be able to predict reasonably well when the two would come apart in order for it to be rational for her to respond to this in her beliefs. A pure internalist version might allow that reward and evidence might even be fully coincident in the environment but allow a trade-off so long as the agent reasonably (and falsely) believes that they will come apart. This is a deep issue about the structure of epistemic normativity that I can't hope to adjudicate here. In lieu of that, I'll proceed using the standard internalist version - that is, by arguing that these conditions both hold of the environment and

can usually be tracked by agents in an internally predictable fashion. I use this conception because it combines the external and the internal requirements, and so by showing that it can be satisfied, I can also establish that the weaker conditions (pure internalist and externalist) can also be satisfied.

There are some key differences between this 'belief bandit' and the standard multi-armed bandit problem. Most significantly, in Vanya's case, the value of the various arms are not independent, since they concern belief (or suspension) about a single issue. In contrast, in the standard bandit, the pay-offs of each arm are independent of one another. This means that in principle, repeated sampling from a single arm in the 'belief bandit' will provide information about the pay-offs of the other arms. However, in practice, nervism and mysticism are quasi-independent: learning one is slightly more likely will not in general make the other slightly less likely. This is because Vanya may not know whether the two hypotheses are genuinely inconsistent or whether they exhaust the possibilities. Further, in the best case scenario, one of these hypotheses as Vanya conceives of them would be approximately correct rather than fully accurate. So while full independence is clearly violated, there may be sufficient quasi-independence for information pertaining to one option not to be equally informative about another option. As I'll discuss below, the relationship of the framework hypothesis to other sub-questions also adds to quasi-independence.

4.1 Background Conditions

Conditions 1 and 2 are easily satisfied by belief. For condition 1, beliefs lead to the acquisition of evidence through experimentation and imagination. The most obvious case is methodological beliefs – if you believe that particle collision is not a very good method of discovery, this will lead you to conduct different experiments and so receive different evidence than were you to believe differently. This is not a fluke. Because our intervention in the environment and our process of imagination are guided by our beliefs, they will fluctuate as our beliefs fluctuate.

For condition 2, the most straightforward kind of epistemic reward is truth. Of course, we don't always know when our beliefs are true. But in plenty of cases, we find out whether we were correct or incorrect, or at least are able to estimate how likely it is that we are right about a particular proposition. To put this in the most flat-footed way, it's possible to treat this value

just like money: when we act (or believe), we aim to receive some amount of reward – a reward that we sometimes observe (in cases like playing a slot machine, or a direct empirical prediction) but often have immense troubling estimating and verifying (in cases like making a career choice or believing a scientific framework).

However, the value of belief need not reduce to truth. For instance, beliefs may have epistemic value if they amount to knowledge. Since this value is not perfectly luminous to the agent, we can treat it as a 'result' in the same way that we considered truth to be a value resulting from the choice over beliefs. Similarly, an internalist evidentialist might think that the value in belief has to do with justification, and adheres even if the belief is false. The fact that justification is an intrinsic feature of a set of beliefs might look a problem for thinking of justification as a 'result' of belief. But in this context, separating the act (belief change) and the result (justification) just means there is some epistemic distance between attaining the state and attaining the reward: you can know that you're in the former state without knowing you've attained the latter. Likewise, it may be the act of eating ice-cream is inseparable from its intrinsic tastiness value, but that fact is sometimes inaccessible to me and so needs to be estimated and learned for the purposes of planning.

I suspect that the simplifications necessary to treat belief as an iterated problem are of a kind with the simplifications necessary in the case of action – in neither case is there literal repetition, but the relationship between a sequence of similar choice contexts is close enough for the idealization to be useful.

One wrinkle is that actions like instances of pulling a lever are obviously segmented, whereas instances of believing are hard to separate from one another. How often am I in the position to say that I've believed in God seven times? But this feature, while interesting, is not significant for present purposes – what is required by iteration is that the same belief problem arises over time so that the agent can vary her behavior if she wishes. Even though beliefs themselves are not properly segmented, we can categorize the evidential situation as segmented and repeatable just as in the case of action. After all, it sounds less odd to say you re-considered your belief in God seven times, or have had seven episodes of doubt.

5 Imagination and New Hypotheses

In the case of action, the exploration/exploitation trade-off occurs because as time goes on and you become more certain about the best choice, that choice has a high and stable expected reward, but a lower and lower expected payoff in evidence. That is, there's a divergence between expected reward and expected evidential value that is essential to the trade-off. In the case of belief, this means a mismatch between expected forward-looking evidential value, and expected backward-looking fit with current evidence. This is Condition 4, which I'll argue for in this section via a theory of imaginative search for new hypotheses.

A related literature in philosophy of science has argued that there are ways of believing a hypothesis that are good for gathering evidence, both evidence about the truth of the hypothesis in question, and evidence about related hypotheses, though they are not optimal in terms of backward fit with existing evidence (i.e. expected reward, in my terms). Railton [40] and Kitcher [27] raise the possibility that individual scientists being committed to a hypothesis *beyond* what the evidence supports might help the scientific community arrive at truth in the long run, in part by incentivising the right sort of experiments. For instance, Vanya might be more likely to spend time and effort on valuable experiments during the year if he fully believed in nervism, and this might make full belief more advantageous in the long run than suspension of judgment even though suspension would be more fitting on his current evidence. This position looks like the kind of divergence specified in Condition 4.

However, adverting to these cases faces a serious objection in this context: why does Vanya need to actually believe the hypothesis in question? Couldn't he merely act as if it were true, or adopt another attitude such as supposition or acceptance? Determining whether acceptance is just as good as belief for the sake of experimentation would seem to rest on empirical claims about human motivation. Instead, to side-step this issue, I'll present an argument based on the role that belief plays in imagination. Since distinguishing belief from alternative ways of acting as if is essential to this argument, I'll now discuss the distinction between belief and these other attitudes. For the sake of clarity, I'll call this alternative attitude acceptance, though it must just as well be endorsement or supposition.

What does it mean to merely accept instead of believe? Since we need a way to separate belief and acceptance without begging the question in either direction, this question should be

answered in functional terms: how do belief and acceptance behave such that we can determine in which category to place Vanya's exploratory framework attitude? The functional role of any mental state can be divided into two parts: upstream, or how that state is arrived at, maintained and altered, and downstream, how that state is used to direct behavior, thought and communication. Accordingly, we might differentiate belief and acceptance by an upstream or downstream[10] functional profile.

In the upstream aspect, some epistemologists have held that we decide to accept but don't typically decide to believe (e.g. Cohen [13]), and conversely acceptance is often invoked as accepting *for a purpose*, suggesting a deliberative act to achieve an end. We can distinguish two different ways of drawing the line here. The first upstream distinction takes belief to be an attitude that is often formed implicitly or automatically, whereas acceptance is always arrived at by an explicit, deliberative process. On the second, belief is involuntary whereas acceptance is voluntary.

With respect to downstream function, beliefs are used to guide action and thought across contexts and questions, whereas acceptance is only used in a restricted partition of relevant downstream contexts (e.g. Fleischer [16]). In other words, we accept something for a particular purpose or in a limited domain whereas when we believe something, we take it to be true regardless of the context. Relatedly, belief, unlike acceptance, seems to be *epistemically assessable*: we can accept for a purpose even propositions that are false, because acceptance does not imply one's epistemic outlook. This is why we can say, "I don't think it's true, but I'll accept it for the sake of argument", but not "I don't think it's true, but I'll believe it for the sake of argument"[11].

Both the upstream and downstream versions of the distinction could be interpreted as categorical or as articulating two ends of a continuous spectrum. I'll remain neutral on which of these ways of drawing the line is the right one, instead that arguing each functional distinction suggests that some of the epistemic advantages that Vanya would enjoy by really believing the framework proposition would not be accrued if he merely accepted it.

I'll now argue that how we believe constrains our imagination, and that this constraint is not merely psychological but rational. Imagination here means something fairly specific; a mental search process aimed at coming to know new possibilities. In Vanya's case, this could

[10] 'Downstream' in this sense is synonymous with what Millikan [34] calls the consumer-based approach

[11] I owe this suggestion to the editors of *The Philosophical Review*

mean imaging a new hypothesis about the vagus nerve and its connection to the stomach, or entertaining the idea of a world with alternative moral norms. Given that we are not logically omniscient, we need to somehow come to know these alternatives, and this process will involve a kind of construction. That is, following Newell's classic theory of search spaces [36], imaginative search involves building candidate possibilities and evaluating these candidates in an iterated cycle: as we build, we have more to evaluate, and each evaluate guides the subsequent construction process. How we construct this space is crucial to our epistemic success. For example, Koedinger and Anderson [28] model experts and novices in geometry proofs as employing different search spaces, which accounts for differences in errors, response times, and attention to elements of the problem setup. The core idea behind Newell's theory is that imaginative search is not a random or brute force operation – and strategic expertise in search involves not only better evaluation, but also a better understanding of the space itself. Expertly constructed search spaces entail that two agents can both be searching via a random walk process, for instance, but still differ systematically in their success depending on how the space in which the walk is conducted is structured.

There are two important features of these cognitive models that I want to bring out. First, they show us that strategic imaginative search is possible: that is, imaginative search that is sensitive to the agent's evidence as well as her capacities. Second, to the extent that search is tailored in this way, we now have the possibility of self-reinforcing cycles - incorrect or misleading expectations about evidence and/or capacities that lead to suboptimal search. Since search in turn feeds back into expectations, this process will in some cases turn into a genuine cycle: bad expectations leading to bad search leading to more bad expectations. But to understand how serious the threat of cycles is, and how exploration in *belief* in particular might reduce the risk of a cycle, we need to understand more about how beliefs inform imaginative search.

In fact, our beliefs guide imaginative search in several ways. First, they might serve as a starting point - many search processes involve progressively relaxing our current theory in order to come up with neighboring alternatives. For example, I might start looking for new theories of evolution by entertaining minimal variations from my current favored theory. Second, our current beliefs serve as side-constraints, coming in during the building and evaluation process. For example, I might begin imagining one version of a theory of evolution only to re-

alize it could not be possible given my current evidence about breeds of cats and dogs. That is, I might so to speak accidentally run into my background belief about cat breeds in the process of entertaining a seemingly independent proposition. Finally, my beliefs allow me to estimate the costs of imagining a particular option and the expected value associated with this cognitive exercise. For example, I might believe that coming up with new geometric theories is beyond my mathematical capabilities based on my views of my own (in)competence.

These three roles for belief in imaginative search distinguish the connection between belief and imagination from that between belief and experimentation as explored by Kitcher and Railton. In experimentation, we use our beliefs or other mental states to design an intervention in the world, and then the world gives us back some information - at least when things go well. But in imagination, the role of the world is played by our own internal model. That is, we have only ourselves to tell us whether some imagined construction is really possible; there is no feedback from the world that allows us to make this determination.

Let's make this analogy more explicit, since avoiding the deadlock over experimentation requires that the role of belief in experimentation and in imaginative search genuinely differ. For the sake of the analogy, I'll describe the form of imagination most similar to experimentation: mental simulation. But note that mental simulation, if not always itself imaginative search, is part of many imaginative search processes. For example, simulating the trajectory of a bullet might be part of searching for hypotheses about who committed a crime.

Now for the example. Suppose you want to know where a ball will go if it is kicked off of a ledge at a certain angle. One way to answer this question would be to play around with the ball and perform one or more small experiments. In this case, the knowledge you end up with is a product of your mental states as well as input from the environment. More specifically, your mental states guided you in setting up the experiment and in interpreting the results, whereas the environment provided you with data concerning the trajectory of the ball and its final location. Now suppose that you went about answering this question through imagination: rather than actually kick the ball, you went through a series of projections of where the ball would go (just like the experiment, this might consist in one 'trial' or multiple). In this case, your mental states have the same roles as in the first case - they led you to set up the mental experiment, and guide you in interpreting its result. But there is an additional, and critical,

role in the imaginative case that was not present in the case of experimentation. Your mental states also take the place of the environment in telling you the trajectory of the ball and its final position. You imagine the ball as having a spherical shape, as not violating the laws of gravity, as having an approximate mass, the weather as not interfering with the kick, and so on. These are what I call side-constraints.

Here's the key lesson of our example. We've seen that in experimentation, many if not all side-constraints can come from the environment, whereas in imagination, side-constraints must be based on background mental states. Now in what follows, I'll make the further argument that while these background mental states can includes suppositions, acceptances, hopes, and so on, there is a special role for background *beliefs* in generating side-constraints in the activity of imaginative search.

Let's now consider how Vanya will imagine during the academic year and during the summers. Instead of ruminating on the central question of nervism or mysticism, he will likely spend a lot of his time imagining smaller questions within these two frameworks: the connection between the brain and the stomach, for instance, or the nature of evil.

We'll start with the first upstream view of acceptance, where acceptance is deliberate whereas belief may be automatic. In imagination, this difference manifests in how a scene is filled in. When Vanya really believes in nervism, he will automatically populate an imaginative scene based on this background belief. This process cannot be deliberative for computational reasons; it must be done quickly, and in parallel as opposed to serially. To see the psychological manifestation of this, consider the following riddle (discussed by Bar-Hillel et al, [7]):

> An accountant says: "That attorney is my brother", and that is true – they really do have the same parents. Yet that attorney denies having any brothers – and that is also true! How is that possible?

The answer, which most people do not discover[12], is that the accountant is a woman. I take these stumpers to illustrate the involuntary nature of side-constraints in imagination. Most people imagine the accountant as a man, but not because they decided to - in a separate imagination task, 71% of participants reported imagining an accountant as male. This does not reflect statistical frequencies, but rather that a male accountant is something like a prototypical accountant;

[12]Over two studies, between 35-48% of participants solved this riddle

in a related experiment, a significant majority of participants reported imagining an Italian as male, even though they presumably would expect Italians to be statistically half female. If participants were capable of deciding how to fill in the imagined scene, these riddles wouldn't work - at least after the first time. On the contrary, being exposed to this particular stumper will likely not help you solve the next stumper (if you'd like to try, another one from Bar-Hillel et al. is presented here in a footnote[13]).

In addition to a lack of explicit thought, this example shows us that some uses of attitudes to guide imagination must also be involuntary, that is, they must fit the second upstream feature of belief. If we could fill out an imaginative scene under voluntary control, it would be possible to solve these stumpers easily once you became aware of how they work by playing with your assumptions. You could decide to fill in the details of the scene one by one, or to only imagine what was literally described. But this strategy cannot be followed, and so it would seem that at least this way of using side-constraints to fill in a scene is not typically under voluntary control. This should not be surprising: filling out the scene is done quickly and voluntary control would make the process cumbersome – as well as distracting attention from other goals.

Of course, the stumpers are hard for most humans, but why think that they show us something about imaginative search in general, or for more sophisticated agents? After all, the specific stumper I provided evidences a kind of gender prejudice, not a rational informing of imagination by belief. I think the best interpretation of Bar-Hillel's results, combined with Newell's search space framework, is that imagination will often involve hard to locate and automatic background assumptions even for highly sophisticated agents, since the function from background knowledge to search space structure is complicated enough that psychologists have struggled to come up with plausible candidates, even at a high level of abstraction ([42]). While Koedinger & Anderson, for instance, provide convincing evidence that geometry expertise changes search space structure in a rational way, they do not (and presumably could not) articulate how. But if this function is that complex, then agents who are more sophisticated than we are would likely also be subject to a certain sort of stumper, because the stumper is just a way of exploiting the opacity of the function from background knowledge to imaginative

[13]Individual bus rides cost one dollar each. A card good for five rides costs five dollars. A first-time passenger boards the bus alone and hands the driver five dollars, without saying a word. Yet the driver immediately realizes, for sure, that the passenger wants the card, rather than a single ride and change. How is that possible? (answer at the end of the paper)

scene. So nothing is directly established by our succeptibility to stumpers. However, the normative idea of strategic imaginative search and an assumption that the function from background knowledge to some features of the search space (such as its structure) will remain complex even for thinkers with greater cognitive capacities, we are now in a position to see stumpers as something other than a human caprice. Instead, they are an expected, though perhaps unfortunate, consequence of a rational activity. In fact, this is essentially what Bar-Hillel herself suggests, though she is more interested in what they reveal about linguistic conventions.

Note that this argument does not necessarily generalize beyond side-constraints. Let's imagine that Vanya is entertaining a few different possibilities about the structure of the vagus nerve. He accepts a series of possibilities in turn, and from that initial point of acceptance, goes on to spell out for himself how things would be if the nerve enervated the digestive system in some particular way. While the first instance of acceptance was deliberate and voluntary, it seems reasonable that the subsequent states should count as acceptance even though they arose automatically and quickly from the initial acceptance. In some sense, they would inherit their nature as acceptance from that initial point, since we would expect their endorsement to remain contingent on that initial point. This is why when he moves on to the next theory, the previous imagined possibilities can be set aside. But side-constraints are different. As Vanya thinks through each possibility, he uses his background views of the world to fill in the details, to ask himself questions, and to generate answers and results to imagined manipulations. The background beliefs that support this process are not pinned to a particular accepted starting point, but instead are used invariably across different imaginative exercises. In fact, it is this automaticity and ease with which we draw on background beliefs that enables the more limited automaticity of a string of accepted propositions.

To generalize, the attitudes that guide imagination as side constraints should be at least somewhat implicit and involuntary, so that they can be used quickly and automatically to fill out scenes. In Vanya's case, fully believing in nervism would lead him to use nervism as a background theory in quickly and automatically populating imaginative scenes. Were he to merely accept nervism, it could not play this background role. And so when he is entertaining the various sub-hypotheses of nervism, we would expect believing in nervism to give him an epistemic advantage in theory search.

This point is related to a more general Wittgensteinian idea, or at least one attributed to Wittgenstein by Crispin Wright, who in discussing the necessity of methodological propositions writes: " By that I don't mean that one could not investigate (at least some of) the presuppositions involved in a particular case. But in proceeding to such an investigation, one would then be forced to make further presuppositions of the same general kinds. The point concerns essential limitations of cognitive achievement: wherever I achieve warrant for a proposition, I do so courtesy of specific presuppositions — about my own powers, and the prevailing circumstances, and my understanding of the issues involved — for which I will have no specific, earned warrant" [51]. This Wittgensteinian idea is meant to apply to evidence search more generally, and is of course quite controversial [24]. But the context of imaginative search is a much more favorable case for the necessity of reliance on background knowledge. Compare a case where I learn what will happen if a glass falls on the floor by actually knocking over the glass, and a case where I learn the same thing by merely imagining a glass falling. I might have to presuppose quite a lot to update my beliefs in the real glass case, but in the imagined case, there are a whole suite of additional roles for side-constraints. For instance, I need to fill in an imagined path for my hand when knocking over the glass, and a size and shape for the glass. Aronowitz & Lombrozo [4] make the case that simulation (a sub-category of imaginative search) is a way of extracting information from latent, opaque mental models such as the motor model that we all build as we acquire implicit knowledge of how our own bodies and other objects interact. When we learn something from the case of the imagined glass, this is only possible because we rely on these models. This is a more developed version of the intuition I drew from the ball kicking case that compared imagination and experimentation. Unlike in the case of inference, Aronowitz & Lombrozo argue, where we are in a position to understand where our new belief came from, imagination requires a kind of reliance without transparency. Thus the debate over hinge propositions concerns a stronger and more controversial claim than the one defended in this section.

In summary, so far, I have argued that the upstream way of distinguishing belief from acceptance has the consequence that beliefs should have an important role in generating side-constraints. This argument took the form of a dilemma: if the distinguishing feature of belief is that belief is automatic, then beliefs are needed to fill in side-constraints quickly and in par-

allel so that we can construct an imaginative scene or theory effectively. If the distinguishing feature of belief is that belief is involuntary, then beliefs are the best candidate for grounding most side-constraints since they don't require a cascade of choices but can be accessed stably and consistently. These arguments are only inferences to the best explanation and rely to some extent on data about the way we actually imagine. As such, rather than proving that belief's upstream role requires that belief be used in imaginative search, I have merely established that there is a reasonable connection between what kinds of states function efficiently for imaginative search and the two candidate upstream features that differentiate belief from acceptance.

The two downstream ways of distinguishing belief take belief to be less contextually constrained than acceptance, or more epistemically assessable. But thinking about the way side-constraints function also makes these features a decisive factor in successful imagining. Let's start with contextual contraint. We often learn from imagination when we see that two things we took to be unrelated are actually related. For instance, Vanya might be sitting at the dinner table and ruminating on the way his mood is connected to the rumbling of his belly. Suddenly, his theory of the vagus nerve might occur to him - on his current nerve model, the stomach is not connected to the brain directly so this response should not be possible! If Vanya were to merely accept nervism, he would be accepting it just in the context of his academic project, or in some other limited context. Consequently, episodes of imagination outside that context would not draw on this background belief. Overall, this might be perfectly fine, but it would reduce his ability to have epiphanies of a certain kind that involve connecting seemingly unrelated ideas.

One might object that Vanya does indeed have a contextually constrained series of attitudes: after all, he accepts nervism in the academic context, and denies it in the context of summer in the dacha. Is there any meaningful difference between this kind of switch, and the kind of switch that we perform when we accept a proposition in the context of a five-minute argument? Could this difference in degree of duration be enough to entail two different attitudes? From a computational perspective, five-minute partitions require careful upkeep and online monitoring to keep the accepted proposition within its proper bounds. Let's assume that there is a fixed amount of cognitive effort necessary to make each shift, in tracking the context as the boundary is approached as well as the operations required to actually shift. As we make the interval between contexts longer and longer, monitoring is less and less necessary in each moment of

cognition, and the attitude held in the meantime becomes closer and closer to a truly unconditional belief. Given all this, it seems implausible that the mental attitude inhabited by Vanya in the depths of the academic year is functionally any different from the one he would have held were he to always believe in nervism with no switches at all. Here, I will appeal to Marx's line in Capital (a paraphrase of Hegel): "merely quantitative differences beyond a certain point pass into qualitative changes".

Turning to epistemic assessability, we can give a very similar argument. Epistemic assessability reflects the fact that a belief encodes the agent's outlook – it can have no asterisk that allows us to avoid questions about why we believe, and responsibility for believing. The agent doesn't distance herself from her beliefs the way she can from her acceptances. Reviewing the motivating case for epistemic assessability reveals that lack of assessability makes contextual constraint conceptually possible, and under some conditions permissible: after all, if I can say "I don't think it's true, but I'll accept it for now", I have just described the opening of a restricted context in which my acceptance will apply. This connection is not so much causal as conceptual: it does not show why contextual constraint happens to arise, but only how it is not a contradiction with the very nature of the attitude in question. But we can conclude from this that contextual constraint is licensed by lack of assessability and restricted (or even eliminated[14]) by assessability. So if assessability distinguishes belief and acceptance, we can refer back to the arguments I've just presented with respect to contextual constraint, except now we can see that belief not just tends to cause different consequences but makes different demands on the believer.

To summarize. Imaginative search is a way of learning what is possible, not just a kind of cognitive rehearsal. But it is also strategic, tailored to both background knowledge and cognitive capacities. A common, but of course not ubiquitous, form of novelty in imaginative search comes from connecting information from areas or topics that are not obviously related. For this to be possible, imaginative search must sometimes draw widely and deeply on background information. If the distinguishing feature of belief, from a downstream perspective, is that belief is not contextually constrained whereas acceptance is, then we can conclude that the surprising collision of distant information could be accomplished only by drawing on beliefs.

[14]I hedge here because assessability might only forbid adopting beliefs we know to be false, but not those we suspect are false.

I've dwelled extensively on distinguishing belief from acceptence. But it's worth noting that on several prominent theories of belief, all of this fine-grained argumentation would be entirely unneccesary. For instance, Eric Mandelbaum [32] and Jake Quilty-Dunn [39] have advocated for a psychological realist conception of belief, on which it is very easy for a stored representation to count as a belief. All that is needed is for the representation to be used in the right kind of cognitive system, following a set of psychological belief dynamics or "laws". Mandelbaum, for instance, argues that we believe *everything* we consider, at least at first, by pointing to evidence suggesting we are inclined to draw conclusions from what we merely consider, particularly when put under cognitive load or other pressure. On this kind of view, imagination clearly always recruits beliefs, among other states, because all my stored representations that are taken to be true even occasionally in inference and action are beliefs. Another more permissive view of belief is advanced by Hawthorne et al [20]: the authors advocate for understanding belief as a "weak" state that could be merely taking something to be probable, arguing that several forms of linguistic data such as Moore-paradoxical sentences are best explained by the weak theory of belief in combination with a stronger norm for assertion. Unlike Quilty-Dunn and Mandelbaum's view, the weak belief thesis does not directly diffuse the debate about acceptance versus belief, since it's plausible that belief might be weak in a probabilistic sense but still distinguished sharply from acceptance. But some of the objections to my distinction between the two states might seem far less plausible on a weak belief view[15]. One of these objections can be drawn from Jane Friedman's work on belief and inquiry [17]: what I have called belief in nervism, in Vanya's case, is compatible with inquiry about whether nervism is really true, whereas belief should be understood as the attitude we take when an inquiry is closed. This identification of belief with closure will not be true on a weak belief view[16] .

I've discussed three possible functional attributes of acceptance, but another objection might focus not on acceptance but an alternative process of imagination. That is, couldn't Vanya merely *imaginatively inhabit* nervism instead of fully believing in it? This activity might be something like how we enter the world of a novel while engrossed in it, or inhabit the per-

[15] As per my earlier comment, we will have to be careful that on a weak belief view, Vanya's oscillating belief is not so weak so as to be compatible with suspension of judgment, otherwise he would scarcely be oscillating at all and the two choices I've put before him would collapse into one.

[16] On a strong belief view, we might respond to a Friedman-style objection by noting that inquiry about whether nervism is true is closed locally for Vanya almost all the time during his oscillations, it's just that when we zoom out and look at his years spent shifting back and forth, we see that the inquiry is open in some broader sense.

spective of Plato while studying his views. What distinguishes imaginatively inhabiting Plato's theory from believing in it might be a partition of my endorsement to just the imaginative context. But as I've argued, this kind of partitioning can't get us all the epistemic advantages of full belief, which rely on the possibility for unexpected combinations of propositions across different contexts. Compare the scholar who just visits Plato's view with the one who is a true believer. These two scholars are like a traditional actor who gets into character right before she goes on stage, and a method actor who spends months living like her character and loosening the boundaries between her real life and the character's life. I am not suggesting that one of these ways of acting or studying is better than the other, all things considered. Instead, I've argued that each way of believing will come with its own distinct way of imagining. Unlike in the case of experimentation, there is no substitute attitude for really believing that will let you imagine in the same exact way, since imagination depends on belief to fill a wider variety of roles. Analogously, there are some real advantages to being a method actor, and some real costs. Much more would have to be said to articulate when exactly the costs are worth the benefits; my goal here has just been to argue that there are such benefits, and that it would be a bizarre coincidence if the way of believing that was the best fit with current evidence turned out to also balance these imaginative costs and benefits in the optimal way.

This connection tells us that some ways of believing will be particularly good for imagination – a forward-looking advantage that consists in obtaining future evidence and seeing the right things as evidence, among other things. However, there is no reason to think that the best way of believing for this purpose will be the way of believing that is the best fit for the evidence. Since we need to explore this same space of possibilities in every possible world, the best beliefs to be our guide in this process cannot depend entirely on the contingent evidence we happen to have at this particular point in time. Instead. on a more fully developed model of imaginative search, we should expect structural features of sets of beliefs to be diagnostic of imaginative advantage. In the original bandit problem, the more you pulled the same lever, the less you learn from each pull. Likewise, in the case of belief, the more you believe a framework proposition, the less informative it is to use that proposition in your imaginative search. That is, both cases exhibit not just occasional deviation between forward-looking and backward-looking value, but a progressive divergence. In Vanya's case, I've gestured at the idea that toggling between two

incoherent framework beliefs might enable a more productive search process than suspension of judgment. Suspending judgment might have the highest expected immediate reward (i.e. myopic exploitative value) whereas belief oscillation may put us in the best position to learn in the long run (i.e. exploratory value).

6 Taking stock

I've argued that there is indeed a strong enough parallel between belief and action to extend the exploration/exploitation trade-off. This argument hinged on the hypothesis that strategic imaginative search is a significant driver of learning, and that what makes this search process strategic is in part a sensitive to background beliefs. In rough outline, the trade-off applies whenever there is a systematic, forseeable diversion between options that have high estimated myopic value, such as adopting the package of beliefs and other attitudes that fits best with current evidence, and options that put the agent in a position to learn the most, such as jumping between fully immersive sets of beliefs in order to enable the broadest and most effective imaginative search. In the bandit case, the classic problem that exploration avoids is being stuck in a cycle where the current suboptimal option looks good and the agent does not acquire evidence that suggests otherwise despite its availability, thus reinforcing the mistaken expectation and leading to the suboptimal action being repeated. The same problem, I've suggested, comes up in belief – this problem reflects the negative side of strategic search, and calls out for exploratory beliefs as a solution.

However, just like in the case of action, this framework suggests that most of the time, believing what best fits with the evidence is optimal, and that occassional deviation from that policy become more helpful in inverse proportion to the position of the agent in her learning trajectory. Exploration should also be sensitive to the overall level of reward or risk available in the environment. Another consequence of this analogy is that while these general features should modulate the trade-off, the agent should not be picking and choosing when or even how to explore. That is, algorithms that manage the explore-exploit trade-off successfully[17] present a recipe for exploration that does not itself consist in an expected utility calculation: the agent merely adheres to a global rule of occasional random departures. This point strikes

[17] This applies even to non-stochastic exploratory algorithms like deterministic Upper Confidence Bound.

me as the most important contribution of the explore/exploit framework to epistemology. What it provides is not a recipe for how to calculate exploration, and indeed the guidance it provides runs out at specifying a level of exploration for a particular agent at a time and in an environment, and perhaps in a given domain. Indeed my argument drew on the way in which belief, as opposed to acceptance, is not just belief for a purpose, or belief in a context, but a state of commitment that extends across aims and contexts. If this is part of the nature of belief, we should expect an exploratory strategy in belief to be diffuse rather than specific.

In fact, the preceeding discussion has brought out an important way in which the belief problem is *not* analogous to the original bandit problem. When I pull an arm on a slot machine, I can easily envision what will happen: I'll get a result, if this result is at all surprising I'll get new information, and I'll use that information to update my beliefs about the goodness of the arm. The benefits of exploration are clear and direct: if I learn something, it will be as an immediate result of my action, and it is already obvious how I will be able to use that infomation in directing future behavior. That is, exploration has a proximal payoff (new information right away) and a distal payoff (an overall more reliable path to knowledge of the environment), and the former at least is predictable and easy to identify. The belief problem is much messier. Vanya, if he explores, will adopt this oscillating pattern which will at some point and in some complex way alter his relationship to his evidence, allowing him to figure out more about both ways of seeing the world than he would if he had suspended judgment or just stuck with the option that looked best. But the link between his exploratory move (the oscillation) and its proximal payoff (better understanding of the possibility space) is far more complex than in the bandit case, although the relationship between exploratory move and distal payoff is extremely complex (and presumably intractable) in both cases. Further, the disanalogy is even more serious than that: we can see this complexity difference between belief and much more realistic cases of action as well, since when I decide to go to the supermarket, my friend's house, or even law school, I antecedently understand the mapping between possible things that might happen and the kinds of information I would gain. In the case of belief, it's precisely because of the lack of understanding the possibility space that I am able to explore, and yet that lack of understand also impairs my ability to form expectations of information gain through exploration.

What is the consequence of this disanalogy? It provides further support for the diffuse,

non-strategic nature of exploration in belief. Without a reasonably good expectation, we cannot look at ourselves at a critical point in belief formation and carefully decide exactly which propositions to explore for exactly how long. However, this does not undermine the trade-off or make exploration impossible. Unlike, say, a classical consequentialist trade-off where I recognize myself at a decision point with a set of well-defined options, random exploration such as ϵ-greedy is designed to improve outcomes in agents without a costly (and in this case impossible) planning process.

So the very features of the belief problem that make sense of strategic imaginative search also create a distinct form of complexity in the belief "bandit" not found in the action bandit. Exploration is not undermined by the complexity of the belief problem for two reasons: first, even through the complexity, we can still distill regularities such as the ones I have analyzed in the previous section. Second, the complexity of the belief problem reinforces the need for explanation because it increases the seriousness of cycles and local minima. This is because the complexity of belief arises from the fact that each belief could in principle be connected to any other belief, whether through deductive reasoning, induction, analogical reasoning, imaginative search, or other cognitive process. But this very fact means that while a bad cycle of behavior and expectation in the bandit case will just have direct ramifications for my behavior at that slot machine, in the case of belief, a bad cycle of self-reinforcing belief and imagination can infect quite a lot of other beliefs (It's worth noting here that the interplay between "random" and "directed" exploration is an active area of study right now in psychology and computer science [50], though it's an open question how to characterize these two modes and to what extent they might be related). So while belief and action do indeed differ with respect to the complexity of their relationship to evidence, this difference does not undermine the trade-off but to the contrary increase the need for a form of exploration in belief such as ϵ-greedy that can be employed even under severe uncertainty.

7 Objections and alternatives

One objection to my account is that if we accept the rationality of the exploration/exploitation trade-off in action, positing an additional trade-off for belief amounts to two solutions to one problem, where each solution is on its own sufficient. That is, isn't introducing exploration

twice overkill?

There's something undoubtedly correct in this suggestion - agents who introduce arbitrary oscillations, randomness or other exploration behaviors at multiple points face a difficulty in making sure these interventions are consistent. In some situations, introducing exploration at the level of action will be enough to reduce the agent's chance of getting stuck in a local maximum (a place in the 'landscape' of belief packages that is better than anything around it in terms of epistemic value, but not the best possible package). And likewise for imaginative search; if we introduce randomness into the search process itself, that will solve some of the problems of a purely exploitative approach. Note that this move still involves changing the canonical framework for exploration in action, since we would be interested here in the epistemic rationality of actions not their practical rationality.

However, this will not always be the case, and there are benefits to belief exploration which do not carry over to imaginative exploration. Consider how it is that Vanya's beliefs allow him to explore neighboring possibilities. It's not just that he happens to explore theories that are adjacent to his beliefs; these theories are *made more accessible* to him by his beliefs. Because he believes in mysticism, through coordination of actions, imagination and other modes of thought, he's amassed resources to understand that theory and how it might be altered to create new versions. For one not familiar with mysticism in that intimate, thorough-going way, it wouldn't be clear, for instance, that there are two versions of the view, one which takes the mystical state of oneness with God to have content, and one which doesn't. Given this, in order to gain the advantage of the incoherent package by only changing actions, there would need to be a coordinated exploratory change to both external actions and imaginative ones. Changing the underlying beliefs is a natural and effective way of achieving this coordination. In other words, fully inhabiting the framework is necessary for exploring these fine-grained questions about divine experience that bear little to no relation to action. Further, even changing external actions and imagination in a coordinated way would likely be insufficient; part of how belief makes regions of possible space accesible is intrinsic, coming from the fact that believing in something involves entertaining that proposition fully, in a way that seems deeper than other forms of non-doxastic consideration.

Another objection is that my view presupposes epistemic consequentialism. Epistemic

consequentialism is the controversial theory that epistemic rationality reduces to a decision-theoretic problem where truth, accuracy etc. is assigned some kind of utility. While most discussions of epistemic consequentialism to date have been act-consequentialism (e.g. Carr [11], Greaves [19], Berker [9], Ahlstrom-Vij and Dunn [2][18], and Singer [43]), the account I've given in this paper uses expected consequences to justify a general principle of modulating belief policies over the course of inquiry. That is, it is a kind of epistemic rule-consequentialism. Elstein & Jenkins [14] have proposed that a version of epistemic rule-consequentialism avoids some of the worries that face epistemic act-consequentialism, while Firth [15] discusses a series of objections that target epistemic rule-consequentialism in particular. It's worth noting, however, that both Elstein & Jenkins and Firth present versions of epistemic rule-consequentialism far more substantive than what would be required to incorporate exploration: on Elstein & Jenkins' account, the rules would include trusting in the reliability of induction and the senses, and Firth takes the rules in question to depend on particular, contingent statistical facts. On the contrary, exploration would be accommodated by merely adopting evidentialism with a small amount of noise that decreased over time, and this alternation would not depend on anything in particular about the actual empirical world (such as the existence of natural kinds, lack of truth fairies, and so on).

What separates the position I've defended here from the idea that it would be epistemically rational to experiment with hallucinogenic drugs in order to enhance imaginative search? As Elstein & Jenkins note, there are possible worlds with truth fairies and those without, and likewise there are creatures for whom taking hallucinogens would cause a positive learning benefit. Even if a particular agent and world are such that she could successfully experiment with drugs according to a rule, the success of that rule would depend entirely on non-epistemic, empirical factors. On the contrary, the exploration/exploitation trade-off and along with it, strategies like ϵ-greedy that solve it, arise every time any non-logically-omniscient agent faces a member of a large set of learning problems. These learning problems are defined by our four conditions: the agent can bring about rewards with some uncertainty, she will interact with the same or similar environments repeatedly, and repeating the actions that have the highest expected reward will tend to provide less and less learning benefit. As opposed to features of brain chemistry, pharmacology, or truth-bestowing creatures, these conditions are features of the epistemic sit-

[18] See also [1] for further discussion of epistemic consequentialism

uation of an agent. They are structural, in the sense that the same formalism applies widely across agents, contexts, values, and types of acts. In addition to differentiating the present project from other forms of epistemic rule-consequentialism, this difference has consequences for understanding the place of exploration in epistemic rationality: I've argued that the exploration/exploitation trade-off grounds a structural, rather than substantive, feature of rational belief.

However, while my argument appealed to expected consequences, exploration in belief is not incompatible with other theories of epistemic normativity. I'll note a way for an epistemic deontologist to accommodate the rationality of exploratory beliefs, and one way for an epistemic virtue theorist to accommodate it. These are not meant to exhaust the possibilities, but merely demonstrate the flexibility of the account.

Epistemic virtue theory could hold that exploratory belief is the expression of an underlying virtue or skill, for instance open-mindedness. So my account of the trade-off now serves to describe what open-mindedness looks like and how it can be distinguished from other features of epistemic rationality, namely whatever goes into exploitation. On a responsabilist virtue-theoretic picture, open-mindedness might be its own valuable characteristic, whereas on a reliabilist virtue-theoretic picture, the argument I've given in this paper shows how exploration is a reliable practice.

An epistemic deontologist is canonically not interested in justifications based purely in the results of believing in some way. They could allow for exploratory beliefs by appealing to other considerations beyond generating the right results, usually something like conforming to epistemic requirements. These requirements themselves cannot be justified by their results, otherwise we have rule-consequentialism. One non-consequentialist justification for a requirement to explore might be that trying out new beliefs is an intrinsic part of being epistemically responsible. The possibility of getting stuck in a local maximum, just like the possibility of hurting someone with a negligent bit of landscaping, would thus dictate responsible behavior even if the agent were not actually at a local maximum or her garden did not actually hurt anyone. Exploration is not a black-box reliability machine like using a crystal ball; it's a practice that's integrated into and regulated by our other ways of believing, and the account I've given here shows how we are always navigating the exploration/exploitation trade-off as we move

through the process of learning.

Does this mean that all rational believers will explore? I aim to have established a weaker thesis: exploring by believing is sometimes epistemically permissible. This follows if we assume the following:

> **Optimality Thesis**: if S believing according to method M has the optimal expected epistemic outcome, and S knows this, it's epistemically permissible for S to believe according to M.

This principle reflects the intuitive idea that what makes methods of belief formation good is how well they work and/or the degree to which the believer can reasonably expect them to work. I have required here that S know that believing with method M would likely lead to the best outcome rather than justifiably believe that it would in order to make the thesis a special case of a variety of different positions. Since knowledge entails other plausible conditions such as belief, justified belief, truth, having the fact in one's evidence, and so on, I state the principle in terms of knowledge. We can imagine a version of Vanya who satisfies this strong kind of optimality: he's not just an oscillator, but he's also read this paper and knows that oscillation is going to help him out in the long run. Oscillation counts as a method of believing because it's a kind of policy rather than a package of beliefs, a policy of seasonal shifts. As it happens, I'm inclined to think that exploration can be permissible for Vanya even if he hasn't read this paper, or even realized that he was oscillating at all, but a defense of that would go far beyond the minimal thesis I aim to establish here: that exploration is sometimes epistemically permissible.

This thesis is controversial since it allows methods of self-fulfilling belief to establish permissibility. For instance, if my belief that I will succeed in general is part of what makes it likely for me to succeed (by, say, increasing my confidence and thus my performance), the optimality thesis tells us that it's permissible for me to believe that I will succeed. Objections to this result are often motivated by evidentialism, roughly holding that self-fulfilling beliefs are not based on evidence in the proper way and so are epistemically impermissible (see [49]).

Exploration in belief shares a feature with self-fulfilling belief: both ways of believing use the (expected) results of believing in order to justify believing in the first place. This is why both are permissible under the Optimality Thesis. But the two cases are different in the following way: self-fulfilling beliefs make themselves rational by making the proposition under consider-

ation true. They are only rational *once they are believed.* On the contrary, exploratory ways of believing do not typically make any change to the truth of the propositions under consideration, and their rationality is in no way dependent on making any such changes. They are permissible because they lead predictably to good epistemic consequences, but in what we might call the standard way. The weirdness of self-fulfilling beliefs is this non-standard, non-ratifiable way, which is not shared with exploratory beliefs. So we can amend the thesis as follows:

> **Optimality Thesis***: if *S* believing according to method *M* has the optimal expected epistemic outcome *in the standard way*, and *S* knows this, it's epistemically permissible for *S* to believe according to *M*.

It's beyond the scope of this paper to reformulate the optimality thesis to reflect this difference properly - specifying the way in which self-fulfilling beliefs are non-standard or circular is a complicated project that requires a comprehensive survey of the variety of ways in which self-fulfillment works. I follow Ahlstrom-Vij and Dunn [2] here in holding that nonetheless, self-fulfilling belief is a quite different problem than the one posed by instrumental justification for belief alone. The standard way might exclude a causal contribution of the belief state itself or it might require ratifiability, to name a few possibilities. I take it to be sufficient in this context to point out that the difference between self-fulfilling and exploratory beliefs is precisely the feature which makes self-fulfilling belief look epistemically questionable.

Another objection is that epistemic exploration is too risky to ever be rational. Unlike ordering the wrong flavor of ice-cream, the damage associated with believing incorrectly may not be limited to a few minutes of bad taste. There is something right in this objection. The way beliefs are intertwined with one another and with other elements of our thought and action makes one bad belief potentially extremely harmful. However, this cuts in the other direction as well; being stuck in a local maximum in the epistemic landscape is also potentially incredibly damaging. That is, eating a meal that's not optimal but is perfectly satisfactory is not so bad. Having a belief that is not optimal but is reasonably accurate could be a disaster. Given this symmetry of risks, the desire to avoid epistemic disasters cannot motivate pure exploitation.

On the account I've defended here, the rational way to believe may well involve some randomness or noise. This raises a final objection: isn't there something wrong with believing at random? This problem is due for more serious discussion than could be offered in the context

of this paper, given the widespread benefits of stochasticity: for instance, see [22] for a formal proof that an agent with a little randomness built in almost always outperforms one that uses a more standard algorithm for approximating rational choice. We might take issue with noisy beliefs in two ways: first, a noisy belief might not seem fully attributable to the believer, and second, realizing that our beliefs are noisy might lead to problematic instability. In fact, both problems can be dealt with using the same argumentative strategy, an appeal to the non-randomness of the more general policy behind the individual belief. I model this move on Ruth Millikan's [35] appeal to faculties rather than single mental attitudes. In the first case, if we're worried that I can't take credit for the success of a noisy belief, this appeal consists in preserving the agent as fully creditable author of the belief-formation policy, which itself is neither arbitrary nor stochastic. The second worry is that when I realize my own belief is arbitrary, I might naturally be thrown into doubt about it – can I really be rational in believing that p while fully understanding that only chance explains why I did not believe $\neg p$ instead? (a more nuanced version of this thought forms part of the motivation in [29].) Here as well, appealing to the non-arbitrariness of the policy goes some of the way towards dissolving this objection. After all, even if on a traditional picture, my beliefs are never themselves random, there will be a fine-grained level of detail of implementation that will presumably be random – or at least rationally arbitrary.

What's fundamentally at issue here is where the right level lies in terms of rational determination. The exploration/exploitation idea is that while in general, we should believe exactly what is best supported by our current evidence, this policy is improved by some trajectory-sensitive addition of exploration (whether noisy or deterministic) to make sure we don't get stuck in a suboptimal loop, limited by our own imaginative processing. Thus the policy it ultimately recommends is mostly but not entirely *decomposable*. That is, let's say we were to take each belief problem one-by-one and ask what the optimal way of believing would be, and then string these recommendations together. On a classic evidentialist picture, this procedure of building up from smaller pieces would recommend exactly what would seem optimal when we approach the entire life-span belief problem as a whole. The exploratory belief policy does not have this attribute, though the built-up policy and the life-span policy are not dramatically different either. This divergence between levels may seem unsettling, but as I've argued, denying

it would mean ignoring a systematic, structural way in which the beginning of inquiry rewards exploration.

8 Conclusion

Our country song asked: "how am I ever gonna get to be old and wise, if I ain't ever young and crazy?". In this paper, I've argued that this same line of thought applies to belief. In the beginning of inquiry, we often should believe in order to explore rather than to exploit, but as inquiry progresses, we should drift towards maximizing evidential value[19]. This is a feature shared between action and belief, and exploits the rational connection between belief and imagination.

An implication is that just as in the practical case where reward variability modulated the trade-off, this analysis of belief gives us room to make a parallel move. Epistemic pay-offs surely vary, and often in a predictable way. I need the right theory more urgently when I'm starting to build my machine or about to go on an expedition. At other times, such as idle inquiry, preliminary stages, or even after the plans for the machine are all in place, the stakes are lower. The framework I've put forward would allow us to say that the epistemically rational behavior depends on the pay-off - and tends toward exploitation in the high risk case and exploration in the low risk case.

In some sense, what I've said here is reminiscent of talk that motivates moving away from belief towards acceptance and other belief-like states. However, by demonstrating a symmetric trade-off in the case of action, I hope to have pushed back against this project. If the exploration/exploitation trade-off is a ubiquitous feature of goal-oriented rationality, then rather than classifying exploratory belief-like states as forming a separate category, we should expect the trade-off to occur over states of a single type. Further, by treating the phenomenon as a trade-off in the rationalization of a single state (i.e. belief), my theory has an advantage in terms of parsimony and strength . In other words, my opponent must explain how beliefs and acceptances combine in regulating behavior during exploration, and this may be a difficult task.

My view is also more flexible in describing the gradient of rational grounds as a modulation of the trade-off, since any mixture of rational grounds for a single proposition in an acceptance-

[19] A lingering issue of scale: does the beginning of inquiry mean something like childhood [18], or something more like the beginning of opening more specific research questions through the week or year?

based theory can only be described by the unfortunate scheme $X\%$ acceptance, $1-X\%$ belief. In other words, it's hard to imagine what it would mean to half-believe and half-accept something, whereas it's easy to see what it means to have a belief that results from being 50% or even 21.87% exploratory, since the trade-off can be continuously modulated through the process of learning. The trade-off I have proposed is naturally graded in a way that matches the underlying normative fact that our circumstances give us reason to explore to varying degrees, shifting over time.

More generally, the choice between acceptance and belief as the states at stake here rests on what we think belief is *for*. On one view, belief is the state that we use in inquiry: it guides us in performing experiments, and in dreaming up new theories. At the same time, belief is the state that most tightly tracks what we hold to be true. If these are both part of the picture of what belief does, then we should not choose a normative framework that starkly separates belief from experimentation and imagination. Instead, we should recognize that having one attitude tied both to modeling the world in response to evidence and to building a basis for future learning will lead to complex and important trade-offs[20].

References

[1] K. Ahlström and J. Dunn, *Epistemic consequentialism*, Oxford University Press, 2018.

[2] Kristoffer Ahlstrom-Vij and Jeffrey Dunn, *A defence of epistemic consequentialism*, The Philosophical Quarterly **64** (2014), no. 257, 541–551.

[3] Arif Ahmed and Bernhard Salow, *Don't look now*, The British Journal for the Philosophy of Science (2017).

[4] Sara Aronowitz and Tania Lombrozo, *Learning through simulation*, Philosophers' Imprint (forthcoming).

[5] Robert Audi, *Doxastic voluntarism and the ethics of belief*, Knowledge, truth, and duty (2001), 93–111.

[6] Peter Auer, *Using confidence bounds for exploitation-exploration trade-offs*, The Journal of Machine Learning Research **3** (2003), 397–422.

[7] Maya Bar-Hillel, Tom Noah, and Shane Frederick, *Learning psychology from riddles: The case of stumpers.*, Judgment & Decision Making **13** (2018), no. 1.

[8] Gordon Belot, *Bayesian orgulity*, Philosophy of Science **80** (2013), no. 4, 483–503.

[9] Selim Berker, *The rejection of epistemic consequentialism*, Philosophical Issues **23** (2013), 363–387.

[20] The solution to the bus stumper: the passenger paid with five one-dollar bills

[10] Jennifer Rose Carr, *Don't stop believing*, Canadian Journal of Philosophy **45** (2015), no. 5-6, 744–766.

[11] _____, *Epistemic utility theory and the aim of belief*, Philosophy and Phenomenological Research **95** (2017), no. 3, 511–534.

[12] William Kingdon Clifford, *The ethics of belief and other essays*, Prometheus Books, 1999.

[13] L. Jonathan Cohen, *An essay on belief and acceptance*, New York: Clarendon Press, 1992.

[14] D Elstein and CI Jenkins, *The truth fairy and the indirect epistemic consequentialist*, Epistemic entitlement. Oxford University Press. Final draft https://www. carriejenkins. net/papers/. Accessed **7** (2017).

[15] Roderick Firth, *Epistemic merit, intrinsic and instrumental*, Proceedings and Addresses of the American Philosophical Association **55** (1981), no. 1, 5–23.

[16] Will Fleisher, *Rational endorsement*, Philosophical Studies **175** (2018), no. 10, 2649–2675.

[17] Jane Friedman, *Inquiry and belief*, Noûs **53** (2019), no. 2, 296–315.

[18] Alison Gopnik, Shaun O'Grady, Christopher G Lucas, Thomas L Griffiths, Adrienne Wente, Sophie Bridgers, Rosie Aboody, Hoki Fung, and Ronald E Dahl, *Changes in cognitive flexibility and hypothesis search across human life history from childhood to adolescence to adulthood*, Proceedings of the National Academy of Sciences **114** (2017), no. 30, 7892–7899.

[19] Hilary Greaves, *Epistemic decision theory*, Mind **122** (2013), no. 488, 915–952.

[20] John Hawthorne, Daniel Rothschild, and Levi Spectre, *Belief is weak*, Philosophical Studies **173** (2016), no. 5, 1393–1404.

[21] Simon M Huttegger, *Bayesian convergence to the truth and the metaphysics of possible worlds*, Philosophy of Science **82** (2015), no. 4, 587–601.

[22] _____, *The probabilistic foundations of rational learning*, Cambridge University Press, 2017.

[23] William James, *The will to believe and other essays in popular philosophy*, vol. 6, Harvard University Press, 1979.

[24] Carrie S Jenkins, *Entitlement and rationality*, Synthese **157** (2007), no. 1, 25–45.

[25] Nan Jiang, Alex Kulesza, Satinder Singh, and Richard Lewis, *The dependence of effective planning horizon on model accuracy*, (2015).

[26] Kevin T Kelly, *Ockham's razor, empirical complexity, and truth-finding efficiency*, Theoretical Computer Science **383** (2007), no. 2-3, 270–289.

[27] Philip Kitcher, *Theories, theorists and theoretical change*, The Philosophical Review **87** (1978), no. 4, 519–547.

[28] Kenneth R Koedinger and John R Anderson, *Abstract planning and perceptual chunks: Elements of expertise in geometry*, Cognitive Science **14** (1990), no. 4, 511–550.

[29] Jason Konek, *Probabilistic knowledge and cognitive ability*, Philosophical Review **125** (2016), no. 4, 509–587.

[30] John R Krebs, Alejandro Kacelnik, and Peter Taylor, *Test of optimal sampling by foraging great tits*, Nature **275** (1978), no. 5675, 27–31.

[31] Aditya Mahajan and Demosthenis Teneketzis, *Multi-armed bandit problems*, Foundations and Applications of Sensor Management, Springer, 2008, pp. 121–151.

[32] Eric Mandelbaum, *Thinking is believing*, Inquiry **57** (2014), no. 1, 55–96.

[33] Conor Mayo-Wilson, Kevin JS Zollman, and David Danks, *The independence thesis: When individual and social epistemology diverge*, Philosophy of Science **78** (2011), no. 4, 653–677.

[34] Ruth Garrett Millikan, *Language, thought, and other biological categories: New foundations for realism*, MIT press, 1984.

[35] ———, *Biosemantics*, The Journal of Philosophy **86** (1989), no. 6, 281–297.

[36] Allen Newell, *Unified theories of cognition*, Harvard University Press, 1994.

[37] Cailin O'Connor and Justin Bruner, *Dynamics and diversity in epistemic communities*, Erkenntnis **84** (2019), no. 1, 101–119.

[38] William H Press, *Bandit solutions provide unified ethical models for randomized clinical trials and comparative effectiveness research*, Proceedings of the National Academy of Sciences **106** (2009), no. 52, 22387–22392.

[39] Jake Quilty-Dunn and Eric Mandelbaum, *Against dispositionalism: belief in cognitive science*, Philosophical Studies **175** (2018), no. 9, 2353–2372.

[40] Peter Railton, *Truth, reason, and the regulation of belief*, Philosophical Issues **5** (1994), 71–93.

[41] Michael Rothschild, *A two-armed bandit theory of market pricing*, Journal of Economic Theory **9** (1974), no. 2, 185–202.

[42] Laura Schulz, *Finding new facts; thinking new thoughts*, Advances in child development and behavior, vol. 43, Elsevier, 2012, pp. 269–294.

[43] Daniel J Singer, *How to be an epistemic consequentialist*, The Philosophical Quarterly **68** (2018), no. 272, 580–602.

[44] Chandra Sripada, *Imaginative guidance: A mind forever wandering*, Homo Prospectus (2016), 103.

[45] Richard S Sutton and Andrew G Barto, *Reinforcement learning: An introduction*, vol. 1, MIT press Cambridge, 1998.

[46] Johanna Thoma, *The epistemic division of labor revisited*, Philosophy of Science **82** (2015), no. 3, 454–472.

[47] Michael G Titelbaum, *Continuing on*, Canadian Journal of Philosophy **45** (2015), no. 5-6, 670–691.

[48] Stefano Tracà and Cynthia Rudin, *Regulating greed over time*, arXiv preprint arXiv:1505.05629 (2015).

[49] J. David Velleman, *Epistemic freedom*, Pacific Philosophical Quarterly **70** (1989), no. 1, 73–97.

[50] Robert Wilson, Siyu Wang, Hashem Sadeghiyeh, and Jonathan D Cohen, *Deep exploration as a unifying account of explore-exploit behavior*, (2020).

[51] Crispin Wright, *Scepticism, certainty, Moore and Wittgenstein*, Wittgenstein's Lasting Significance, Routledge, 2004, pp. 241–261.

[52] Wojciech K Zajkowski, Malgorzata Kossut, and Robert C Wilson, *A causal role for right frontopolar cortex in directed, but not random, exploration*, Elife **6** (2017), e27430.