Exploring by Believing

Sara Aronowitz

June 16, 2016

Introduction

In many decision-making scenarios, we can observe a trade-off between choosing the action that maximizes expected value, or the action most likely to result in learning something new: the **exploration-exploitation trade-off**[1]. For instance, if you were choosing between ordering your favorite ice cream flavor or trying a new one. Exploiting means picking the option most likely, on your estimation, 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 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?

I 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. These two connections are valuable insofar as they lead to better beliefs in the future: they generate forward-looking reasons to believe. While past work in philosophy of science has delineated what seem to be exploration-type reasons for belief based on the connection between belief and experimentation, the connection with imaginative search is relatively under-theorized. So in this paper, I use the connection between belief and imagination to support the following claims:

1. Rational imaginative search generates forward-looking reasons for belief for agents who are not logically omniscient.

2. If there are forward-looking reasons for belief, packages of beliefs can be evaluated in terms of the exploration-exploitation trade-off.

3. It is sometimes rational to trade exploitative epistemic value for exploratory epistemic value, most typically in the beginning of inquiry.

If successful, this paper presents an instance of what I call inquiry-reasons - reasons for belief which are purely epistemic but depend on the place of the agent in the process of inquiry.

While past work has focused on situationally driven trade-offs in epistemic value that are arbitrary and often fantastical[3], this paper posits a systematic and universal mechanism.

[1] This paper builds off of joint work with Sydney Keough, and I am deeply indebted to her. Thanks to Meg Feeney, Will Fleischer, Hilary Greaves, Reza Hadisi, Victor Kumar, Peter Railton, Eli Shupe and Chandra Sripada for helpful comments.

Here's the structure. In § 1, I describe the exploration/exploitation trade-off in action, and give some background. In § 2, I lay out two examples of exploration in belief, and in § 3 I argue that the trade-off encodes a rational requirement on belief. I lay out a sketch of a theory of imaginative search in § 4, and show that on that theory, and on any theory where belief rationally constrains imagination to a sufficient degree, the exploration/exploitation trade-off falls out naturally.

1 The Exploration/Exploitation Trade-Off

In § 1.1, I'll explain the trade-off through a simplified version of a classic setup in the literature: the multi-armed bandit. Readers familiar with the trade-off can skip to § 1.2.

1.1 An Example

First, some setup. Let's assume a simple expected utility framework. We have some agent, who has probabilities over various outcomes and multiplies these with the corresponding utilities to generate expected utilities. Canonically, in expected utility, 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 *myopic* in this way.

Now here's the problem she faces. She can choose to play at one of five slot machine arms $i - m$. After each play, she may continue at the same arm, or switch to a different arm - in other words, this is a sequential choice problem. Each arm produces stochastic rewards distributed around a fixed unknown bias [2]. Let's say she starts with the following estimations of the biases, where a higher bias means a higher probability of a valuable outcome: $b_i = .5, b_j = .2, b_k = .1, b_l = .9, b_m = .3$. 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 estimated bias. She would start by choosing arm l. After she plays l, she'll get some information. Let's say that the true bias of l is .8, and the outcomes in the short term reflect that bias fairly faithfully. So by using this method, she will continue to choose l over and over again, because its estimated bias will never drop below that of arm i, which has the next highest estimate. This is the method recommended by her myopic expected utility rule; I'll call it the myopic exploitation method. It's exploitative because it always does what is best according to current expectations - where A is the act which maximizes expected utility, myopic exploitation always requires her to do A.

How good is the myopic exploitation method? Exactly as good as our agent's expectations. If she's right initially that l is the best arm, she'll do fine. However, if she's wrong, and for instance m actually has a bias of .9, her total reward will not be as high as it could have been. Myopic exploitation has a significant risk of getting her stuck (or ending up in a local minimum) - once she's in the situation described above, she'll never stop making the same suboptimal choice.

A very simple way of allowing for exploration to an exploitative decision strategy (where A is the act with highest exploitative value) is to add this rule: at every decision

[2] Multi-armed bandit problems tend to have looser assumptions around bias, for instance that the reward state evolves according to some unknown Markovian function [6]

point, choose a random arm other than A with probability ϵ, choose A with probability $1-\epsilon$. This is called an ϵ-greedy strategy. Here, as we increase ϵ, our agent explores more and more, and correspondingly, any decrease in ϵ will lead to an increase in exploration. As ϵ approaches 0, our agent will learn a lot by choosing all the options equally, but her learning will not benefit her, since her knowledge about the options won't effect her behavior at all. As ϵ approaches 1, her behavior will converge to the myopic exploitation rule - she'll always maximize, and never veer off course. 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.

For instance, in the multi-armed bandit case sketched above, an ϵ-greedy strategy with a sufficiently large ϵ will cause our agent to occasionally sample the arms other than l. As she does so, her confidence will rise that arm m is actually better. As we lower ϵ, she'll only choose m, which is the optimal strategy.

As it happens, ϵ-greedy methods approximate optimal solutions to the multi-armed bandit problems in many contexts. The extant optimal solutions involve calculating the Gittens index of each arm, which is roughly the value that we place on continuing to use that arm adjusting for the potential of learning. This is a computationally expensive procedure (relative to Upper Confidence Bound[1] (UCB) and ϵ-greedy) that relies on forward induction [6].

1.2 Extending the Example

You might be thinking that this trade-off only occurs because myopic expectations aren't very good. However, it still shows up for agents who think many steps ahead - for instance in reinforcement learning (RL), a framework for learning in AI. In RL, the agent calculates the value of each progressive step that she might possibly take, multiplied by a discounting factor (this sometimes includes every possible act, or is cut off at a future horizon - see [4] for arguments that employing a horizon may actually be optimal). So the trade-off isn't dependent on myopia. But one thing to note here is that the reward system in RL is somewhat different than in classical expected utility theory due to the intractability of calculating every possible state of the world. In particular, rewards in RL are thought of as inputs from an external system (but usually one internal to the agent) - this means that the value associated with learning a piece of information doesn't fall out automatically the way it would in a complete EU calculation. To encode the value of information, RL systems often encode a direct value for novelty or surprise[10][2]. There is still a reason to explore in RL because there's still a risk of getting stuck in a local minimum. The only difference is that now it's a minimum over whole plans rather than single repeated choices.

Now let's look at what happens to the trade-off in a standard decision theory context where choices are over complete states of the world. Can we map exploitation onto maximizing expected value - and if so, what is exploration? It might at first seem like this won't work. For every exploration-type advantage, there is an expected-utility rational. For instance, I should try each arm because even if there is a small likelihood of finding one better than l, the cumulative long-term reward of switching in that world is quite high. Also, if expected utility maximization is the paradigm of practical rationality, it shouldn't be this easy to find a rationale for violating it. This question goes somewhat outside the scope of the present paper (mainly because it touches on tricky issues about EU in a sequential choice context), but I'll note two ways to deal with it.

One way of incorporating the trade-off here is to allow that exploitation value is something more narrow than expected utility. By making this concession, the trade-off is something at a level below EU that explains how EU is computing. Talking at this level will help explain things like variations in behavior over time. But it muddies the waters a bit because it's not longer clear what exploitation value could be.

Another option is to insist that while zooming out to this higher-order level can accommodate the behaviors I've described in the multi-armed bandit case, there's a trade-off at work here too. That is, exploration is a way of hedging on our expectations, and so it should apply even if these expectations are complex and higher-order. Nearly any optimal EU agent with misleading evidence will benefit from exploring, where that means going off policy. Again, EU is only as good as an agent's expectations. While the simple multi-armed bandit I presented isn't complex enough to display the value of veering off of a full EU strategy, all we need to do is expand it to involve more radical uncertainty (for instance, uncertainty about what the options are, or about where new information might come from). So on this view, exploitation maximizes expected value, exploration puts the agent in a position to learn, and the right combination of the two leads to maximum cumulative actual (not expected value).

Now, you might be worried that this trade-off is a feature of the formalism, absent or not explanatory in the informal context. If you weren't convinced by the ice cream example, 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?". I think this expresses a common sense idea. 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. The modulation of the trade-off over time is what makes it explanatory; young Frankie Ballard should say yes to things that old Frankie Ballard should say no to. The reason for this is just the rationale I gave in the formal case.

Finally, one feature of the E/E trade-off in action will be crucial for belief: the relationship between E/E and time. As I noted at the end of § 1.1, 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, there's a relationship with 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.

Reward changes in the environment also modulate the trade-off. Traca and Rudin [11] 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 than at lower traffic times.

Variations in uncertainty, potential for actions, and total available reward all modulate the exploration-exploitation trade-off in actions. As I now turn to belief, I'll show how each of these factors has a parallel in that case as well.

2 What are Exploratory Beliefs?

To describe the trade-off in belief, I start by assuming that exploitation value for a belief is something like the evidential value of that belief. It represents your best guess about how true that belief is, based on your evidence as well as background assumptions and other rationales. What distinguishes these is that they are narrow - they are reasons to believe P because they contribute to or increase the subjective likelihood that P is true. This maps on nicely to exploitation value, since it means making the most of our current expectations. Note that it's analogous to myopic exploitation value, since it concerns the value of a single belief at a single time.

I'll call "predicted accuracy" the analog of total expected value. Predicted accuracy here will be the agent's best guess, given her current evidential situation, at which of her beliefs are true (or likely, in the case of degrees of belief, which I will mostly leave aside) and more generally, which epistemic resources and experiments are likely to be informative. When we are considering two possible belief states for her to move into, or two books for her to read, we can compare them on this metric.

But isn't learning part of predicted accuracy? A way to see that they diverge is to look at epistemic modesty. Imagine an agent with a set of beliefs and a standard of evaluation for those beliefs and the evidence-gathering they rationalize. This agent judges that her beliefs score higher on her standard than any other beliefs; arguably, this is a necessary condition for a stable doxastic state. However, she also has reason to suspect that while her standard of evaluation is the best one she could have given her evidence, when she gets more evidence, she will almost certainly be required to exchange her current standard for a new one. This modest agent, then, can't possibly be evaluating the prospect of a future, improved standard relative to her current standard - instead, she is capable of separating predicted accuracy from what we could call second-order accuracy. Because of this difference, by occasionally and strategically veering off of the policy recommended by her current standard, we might say that she is attempting to learn despite sacrificing predicted value.

One way of understanding predicted accuracy and evidential value is in terms of expected accuracy in the formal sense, but it could just as well fall out of a range of different theories, including informal ones. But what is exploration value? This section is aimed at answering that question. I'll present two cases of belief where the trade-off structure is preserved: the agent can exchange the current evidential value of their beliefs for an improved likelihood of learning.

2.1 Commitment

One class where we can find exploration-exploitation trade-offs in belief is in cases of diachronic consistency, or commitment. In these cases, the agent continues holding on to their epistemic project for a bit longer than they ought to from an epistemic utility maximizing perspective. The case I'll present is patterned off group-level cases from Kitcher[5] and Railton[8]. In their cases, a scientist has a reason to stick with their belief beyond what is rational for the good of the community - in my case, it will be for the good of her future self.

> **Determined Reem:** Reem has been in grad school for a couple of years, and is starting to see all kinds of problems in her supervisor's presuppositions. She decides she'll put these doubts aside and stick with the project for now;

> after all, if everyone gave up in times of doubt, progress would never be made.

For Reem, fully exploiting would mean dropping her confidence in her supervisor's position - this response is what's warranted by her evidence. However, the learning value of sticking with her the project is higher - she'll put herself in an excellent position to verify key claims by sticking with her belief. It might seem somewhat counter-intuitive that she can explore by sticking with her current beliefs, but think of it as deciding to go on a deep sea expedition - she's committing to trying on these beliefs for longer to really see what it's like.

In other words, Reem's doubts mean that she no longer thinks her belief in the value of the project maximizes evidential value. However she chooses to keep that belief anyways, in order to facilitate the success of the experiments, or because she has higher-order doubts about her own doubts. I'll take up the rationality of this response in § 3.

2.2 Diversity

However, sometimes diachronic consistency can get in the way of learning. Here's Quine on what to do with two fully empirically adequate theories which are incompatible:

> We should indeed recognize the two as equally well warranted. We might even oscillate between them, for the sake of a richer perspective on nature. But we should still limit the ascription of truth to whichever theory formulation we are entertaining at the time, for there is no wider frame of reference. [7]

In this spirit, I think that there are cases where either vacillating between competing, jointly incoherent options or holding on to both at the same time can have an exploration advantage. Consider the following case:

> **Unconflicted Nima**: Nima has a day job in a laboratory and at night studies theology. Before starting this lifestyle, he was deeply unsure about the existence of God. However, he will do best at the laboratory if he's able to totally mentally invest himself in his work, and so he adopts a belief in a scientistic naturalism that makes God's existence unthinkable. At night, he won't be able to stay awake as he reads about God unless he fully believes in God's existence. So, he adopts a belief in God as well.

We could imagine someone like Nima who flips his belief switch after he gets home from work, but instead, Unconflicted Nima somehow manages to hold both beliefs at the same time. We could also imagine a Nima who can't hold both beliefs either at the same time or one after the other over and over again without some psychological pain and existential doubt (call this version 'Conflicted Nima').

Now, since Unconflicted Nima is so divided, we might have trouble even generating a function for his predicted accuracy. However, we know that no proper scoring rule (at least) will assign the highest possible accuracy to a set of beliefs containing a contradiction. This means that by adopting the strategy I've described, Nima is by definition taking a hit in expected accuracy - and this naturally generalized to the informal case. Further, Nima presumably does not plan on being 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 God is one that comes with an entire moral and descriptive framework, it's reasonable to think that either future coherent state will have very different standards of evaluation, and recommend distinct experiments. So Nima is also in a state of meta-uncertainty, which he responds to by moving to a less accurate state (in this case by *all* standards) that improves his prospect of learning.

3 Justifying Doxastic Exploration

Even if there are such agents and situations as in 'Determined Reem' and 'Unconflicted Nima', to demonstrate an exploration-exploitation trade-off, I also need to show that these situations can happen to rational agents, and that those agents can, under certain conditions, be rationally permitted to move along the spectrum. Since I've already shown how these exploratory behaviors can and typically do involve taking a hit in current epistemic value, the difficulty is to show that these responses are ever rationally required and for what kind of agents.

I'll present two possible answers. The first answer applies to agents who are incapable of totally separating motivation from belief, and the second to agents who are not logically omniscient. Note that while these are perhaps deviations from ideal rationality, they are neither obvious ones nor involve straightforward limitations in computational power.

3.1 Belief and Evidence-Gathering

Evidence-gathering by definition opens up the prospect of changing beliefs and evidential standards; when we plan on getting evidence, we plan on changing our beliefs. This feature seems like enough to allow a rational agent to prefer exploration in the practice of getting evidence at least - but can you vary evidence-gathering practices without changing current beliefs, and in particular, without giving up some current predicted accuracy?

Maybe all we need is to accept some helpful propositions rather than change our beliefs. An intuitive response to Nima's case is that the best thing for him to do would be to accept the two incoherent propositions about God's existence, rather than believe them. This is presumably what would be recommended by Van Fraassen [12] among others.

What is meant by acceptance? On one end of the spectrum is the view that acceptance just involves acting as if something were the case without changing any beliefs or making any new inferences. On the other end is the idea that acceptance is descriptively identical to belief, it just falls under different normative standards. The latter in this context is question-begging, since I'm arguing that the normative standards for belief should cover exploration as well. Likewise for the suggestion that acceptance and belief differ only insofar as beliefs are only sensitive to direct, truth-seeking considerations (see Shah and Velleman [9]). So I'll consider two cases, one where acceptance is just acting-as-if, and the other where it involves some doxastic changes but not as far as in belief (perhaps inferences drawn from the accepted proposition are quarantined away from the agents other beliefs, for instance).

In the first case, if Nima accepts the existence and non-existence of God, but keeps his .5 credence, there are two issues. First, it's not clear that it's even possible to act as if God exists and does not exist - acting does not allow for synchronic partitions the way belief does. Second, given some plausible psychological postulates, the advantages that Nima gets from believing will not transfer to acting-as-if. For instance, it was his wholehearted investment in God that allowed him to stay up late studying after work; if he was motivated by his belief, and acceptance is just acting-as-if, then we have nothing to replace that motivation. Treating acceptance as a mere pattern of action by definition does not give us any resources to explain motivation.

On the other hand, let's say acceptance involves some but all of the internal properties of belief. Which features might then separate the two? One idea is that acceptance might be conditional, ready to be taken back when necessary; we might want to make sure to keep track of all the inferential consequences of an accepted proposition. I'll come back to this in the next section.

In general, the connection between belief and evidence-gathering gives us a reason to be willing to take an accuracy hit in order to explore. Agents like us are motivated to gather evidence that we judge to be promising, interesting or fruitful based on our other beliefs. These experiments form epistemic projects, and can be said to rely on a shared basis of belief in which we need to be invested in order to be motivated to gather the relevant evidence. However, even if this motivational structure were not in place, we might expect that rational agents would be constrained in their experimentation by their beliefs. Even if we place acceptance between belief and evidence-gathering in some cases, this relationship will still persist, albeit indirectly, since beliefs will affect acceptance which will affect evidence-gathering. Thus some of our beliefs can still be evaluated based on their consequences for acquiring evidence, though these consequences will be somewhat indirect.

To give up on a *normative* connection between what we believe and what evidence we collect is to invest in an epistemology that starkly separates evidential response from inquiry. This move should be a kind of last resort, since it involves relinquishing a plausible route to explaining why experiments are justified. In the next section, I'll argue that the creation of a hypothesis is a kind of internal inquiry that depends on beliefs but carries consequences for future learning.

3.2 Consideration and New Hypotheses

In this section, imagine that Nima has an unlimited capacity to control and shape his own motivational faculty, and let's add that Van Fraassen was right, and he should accept, rather than believe, the existence and non-existence of God. He'll go ahead and behave as if he doesn't believe in God during the day, and does at night, but he will keep his initial .5 credence.

This Nima, I'll argue, will face a dilemma when it come to constructing new hypotheses. Normally, for an agent who isn't logically omniscient, imagination is one route to forming new theories. In particular, we don't imagine totally random possibilities out of the blue, but we use our current understanding of how things actually are, together with beliefs about what is and isn't possible, to come up with new options. For instance, he might start by visualizing an atom according to his current theory, and sort of play around with the possible shapes (according to his ideas about what shapes can do (innate or otherwise)). By this process, he comes up with a few alternative ideas.

See fig. 1 for a scheme for how this kind of constructive mind-wandering works, and figure fig. 2 describes some of the possible constraints imposed by belief. Nima starts out by considering something consistent with his current theory, and relaxes various assumptions in an incremental way in order to explore the (infinite) space of possible

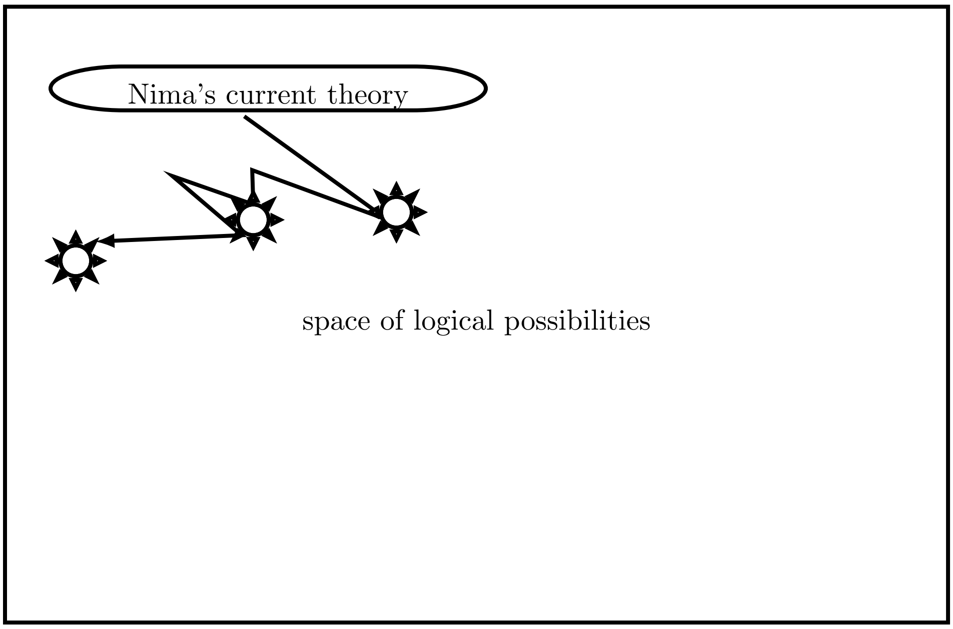


Figure 1: Nima's wandering consideration

hypotheses. The stars represent possibilities which he has fixed on as salient alternatives to the current theory. Nearness here stands for some kind of conceptual similarity - this might be constructed by the search function and hence subjective, or objective relations among the propositions.

Now, the original 'Unconflicted Nima' had two starting points in the space, the belief in God and his other related theological commitments and the more scientific cluster. Some of the epistemic benefit that Nima gets from having the two incoherent theories is that he can explore the space of possibilities in two very different ways that advance both projects. What if his starting points are not beliefs but acceptances or suppositions?

In the case of occasional theorizing, we often start from a supposition rather than a belief. Or when reading fiction, we can start our wandering from the fictional cluster of ideas and move outward, perhaps filling in other details of that imagined world by varying the fictional events ('what if Dr. Zhivago hadn't gotten on that train?'). So couldn't Nima start from acceptance?

Here I think we can allow that while we sometimes begin our wanderings from suppositions, in general there is an important relationship between belief and consideration that doesn't carry over to acceptance. For one, beliefs can guide our wandering not just as starting points but as side-constraints; in this case, we don't need to attend to the constraint in order to use it, in fact in the case of imaginative resistance, philosophers have debated extensively just what some of the common side-constraints might be. Acceptance, in contrast, even in a very belief-y conception, is separated somewhat from other beliefs and inferences, and so won't intrude as much into mind-wandering. Instead, to employ supposition or acceptance in imagination, we'll need to either explicitly trigger the accepted proposition itself or already be 'within the quarantined zone' so to speak. In other words, in imaginative resistance, we have trouble turning off some of our existing beliefs about, say, morality. Acceptance, by definition, is easier to turn on and off than belief, so does not intrude as much into the exercise of imagination,

points of contact between belief and imaginative search

1. Starting points: Imagination often involves starting with our current beliefs about the world, and departing from them incrementally. So in general, we explore neighboring theories first.
2. Side-constraints: Imaginative exercises often involve coming up against previously implicit limits - for instance, in thought experiments when we observe an unanticipated reaction in ourselves that shapes the way we fill in a case.
3. Goals: Imaginative search is a combination of pure random wandering and goal-oriented activity, and the goal-directed end of the continuum depends on beliefs about what's valuable and how to achieve those ends.
4. Stopping rules: For a computationally limited group, or one interested in computational efficiency, it will be necessary to regulate the amount of search, for instance allocating more time to search when the group seems to have reached a plateau or less when current possibilities overwhelm resources to pursue confirmation of those possibilities.
5. Costs and cost analysis: Some theories cost more resources to build than others - another consideration that favors nearby possibilities.

Figure 2: Sketch of possible relations between beliefs and imaginative search

including the use of imagination under discussion - building new theories.

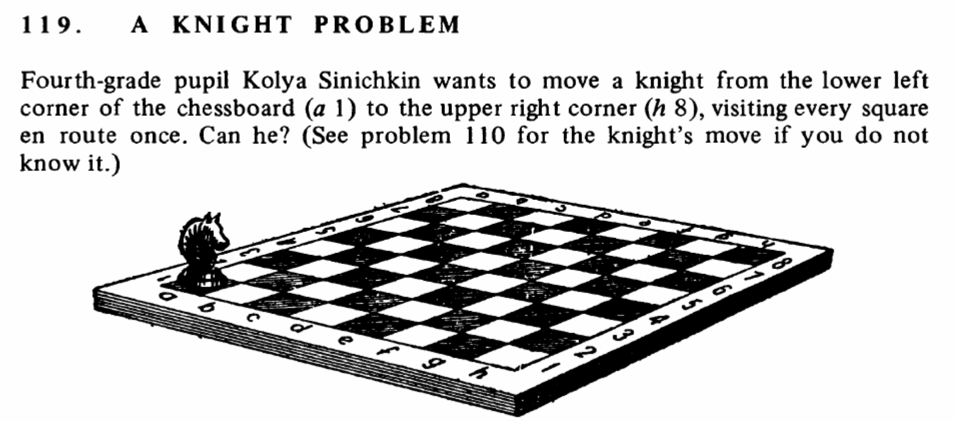
But if I've correctly pointed to a difference between the use of belief and acceptance in imagination, why think that acceptance won't do the job better than belief? In other words, what's so good about the kind of side-constraint that belief generates? To answer this question, I'll now give a sketch of how imaginative search works in a problem-solving scenario.

4 Search Spaces

This section is an elaboration of the suggestion in § 3 about the connection between belief and exploring new hypotheses - though it's not intended to be the only connection between those states. On a more dialectically important note, search spaces are a way of refuting the claim that epistemic value is all about narrow evidential reasons and cannot be forward-looking. These structures are inseparable conglomerates of evidential and forward-looking considerations, and so if we use them in reasoning, we have to let forward-looking reasons in along with them. But first, what is a search space?

A search space is a way of thinking of (a certain kind of) problem solving which goes at least back to Newell. These are, roughly speaking, ways of dividing up the space of possibility. Typically, problem solving involves a tandem process of building and evaluating search spaces - and insofar as these two processes are inseparable, they present a clear case of the interaction between belief and imagination.

A canonical example of a search space is a decision tree for chess. Usually we search through these spaces in practical rationality - the space encompasses possible states which we could bring about, and as we search, we assign values to states. In theoretical



119. A KNIGHT PROBLEM

Fourth-grade pupil Kolya Sinichkin wants to move a knight from the lower left corner of the chessboard (*a* 1) to the upper right corner (*h* 8), visiting every square en route once. Can he? (See problem 110 for the knight's move if you do not know it.)

Figure 3: Note that “once” means once and only once!

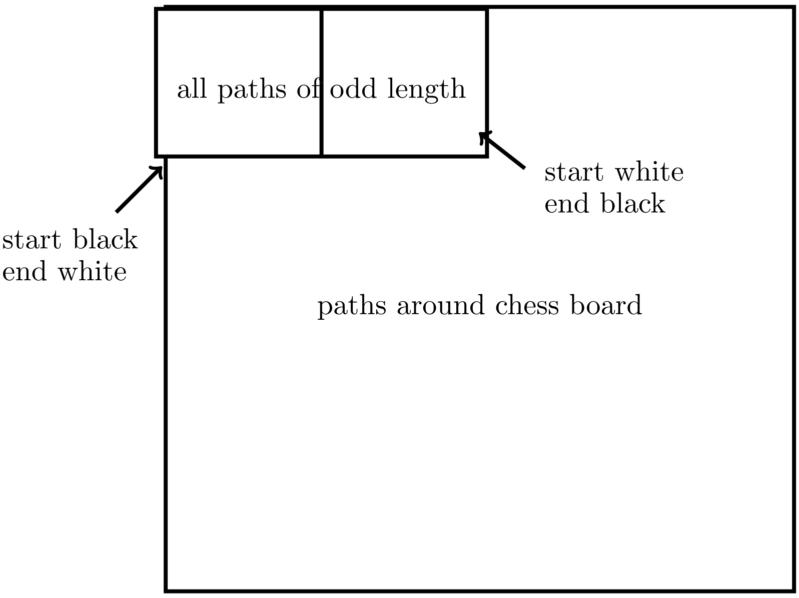


Figure 4: Does this help you solve Figure 2? (nb: not to scale)

rationality, we're trying to delineate a **feasible region**, whose states are epistemically possible. In either case, we can use the search space to answer meta-questions, such as “which proportion of the states have property X?” .

Consider the puzzle in fig. 3. One blunt way of solving it would be to try every possible series of moves, which we can implement as building a tree starting with the first two possible moves and spreading onwards. When could we stop building this tree? Either when we found a path that worked, or when all possible paths had been enumerated.

On the other hand, consider the alternative search space in fig. 4. This space should help you solve the puzzle - since the path across the chessboard touching each square once and only once is 63 moves long, it must start and end on opposite colors. Therefore, it's impossible, since both corners are black.

In effect, you used a very coarse-grained search space to solve this puzzle much faster than would have been possible by using the finer-grained tree representation described earlier. This shows the importance of getting the right space for the right problem.

However, these representations are often involved in solving multiple problems, across time and accumulation of information. As such, a search space can be evaluated in 2 ways:

1. Narrow: This search space will lead us to accurate beliefs in the problem for which it was constructed given the agent's current credences.
2. Broad: This search space will lead us to accurate beliefs by its future use in unrelated problems or when the current problem changes.

Even narrow value can come apart from predicted accuracy because of the way online search works - *beginning to construct a search space is a forward-looking endeavor already.*

What I mean to suggest is that sometimes, we solve problems by choosing between developing different search spaces. Once we choose one kind of space, we begin to build it. As we build it, we evaluate the various states in it according to the problem. This evaluation itself influences the building process - which by necessity influences the evaluation, since it's a function from states currently enumerated in the space to values. Thus, the use of search spaces is an illustration of the general phenomenon in the previous section - the forward-looking and inescapable interplay between beliefs and imagination. Since exploring the space of possibilities through imaginative search will be more useful at the beginning of inquiry (as opposed to the end where the state is already mapped out), this connection will be of evolving epistemic value as the process of learning progresses.

5 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 should believe in order to explore rather than to exploit, but as inquiry progresses, we should drift towards maximizing evidential value. This is a feature shared between action and belief, and exploits the rational connection between belief and imagination.

A further connection 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 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 payoff - and tends toward exploitation in the high risk case and exploration in the low risk case.

In some sense, most of what I've said in this paper is the kind of thing that has motivated the project of moving away from talk about belief to 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 attitudes, 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, my theory has an advantage in terms of parsimony and strength - as well as flexibility in describing the gradient of rational grounds, since any mixture of rational grounds for a single proposition in a two-state theory can only be described by the unfortunate scheme $X\%$ acceptance, $1 - X\%$ belief.

More generally, the choice between acceptance and belief as the attitude at stake here rests on what we think belief is fundamentally about. On one view, belief is the attitude that we use in inquiry - it guides us in performing experiments, and in dreaming up new theories. If this our picture of what belief does, then we should not choose a normative framework that starkly separates belief from experimentation and imagination.

References

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

[2] Nuttapong Chentanez, Andrew G Barto, and Satinder P Singh. “Intrinsically motivated reinforcement learning”. In: *Advances in neural information processing systems*. 2004, pp. 1281–1288.

[3] Hilary Greaves. “Epistemic decision theory”. In: *Mind* 122.488 (2013), pp. 915–952.

[4] Nan Jiang et al. “The Dependence of Effective Planning Horizon on Model Accuracy”. In: (2015).

[5] Philip Kitcher. “Theories, theorists and theoretical change”. In: *The Philosophical Review* 87.4 (1978), pp. 519–547.

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

[7] Willard V Quine. *Theories and things*. Harvard University Press, 1981.

[8] Peter Railton. “Truth, Reason, and the Regulation of Belief”. In: *Philosophical Issues* 5 (1994), pages. ISSN: 15336077. URL: http://www.jstor.org/stable/1522874.

[9] Nishi Shah and J David Velleman. “Doxastic deliberation”. In: *The Philosophical Review* (2005), pp. 497–534.

[10] S. Singh et al. “Intrinsically Motivated Reinforcement Learning: An Evolutionary Perspective”. In: *IEEE Transactions on Autonomous Mental Development* 2.2 (June 2010), pp. 70–82. ISSN: 1943-0604. DOI: 10.1109/TAMD.2010.2051031.

[11] Stefano Tracà and Cynthia Rudin. “Regulating Greed Over Time”. In: *arXiv preprint arXiv:1505.05629* (2015).

[12] Bas C Van Fraassen. *The scientific image*. Oxford University Press, 1980.