**A Planning Theory of Belief**
Sara Aronowitz, University of Toronto

1. What does it mean to hold a belief? Some of our ways of speaking in English suggest that to hold a belief is to have something in your mind: beliefs are things we acquire, defend, recover, and so on (Abelson, 1986). That is, believing is a matter of being in a *state* of having a *thing*. In this paper, I will argue for an alternative: believing is something we do. This is not a new suggestion. For instance, Matthew Boyle (2011) defends a theory of belief as an activity, which he traces back to Aristotle. This paper, however, makes two new contributions: first, I argue for an analogy between belief and planning that fleshes out what it would mean to understand belief as an activity, and second, I aim to show how the resulting view can help sense of a variety of theories in cognitive psychology that suggest cognitive information storage is dynamic and reconstructive.

Imagine you are eating breakfast while planning your route to work. In virtue of what do you count as planning? The most obvious answer is that you are conducting a certain kind of mental activity, a kind of search for the best (or perhaps merely an acceptable) way to get to work. Further, as I’ll argue, this mental activity can’t be reduced to holding in one’s mind a series of plans – it’s possible to be planning without having a plan yet, as when you just start thinking, and plans can also be brought to mind without planning taking place, as in when you implement a plan that was formed earlier on.

The belief-as-planning account draws on these features. Our theory of belief should really be a theory of *believing*, a mental activity that is not reducible to a set of beliefs. By placing this activity at the center of the account, we can now see beliefs, the sets of things, as an abstraction over the activity of believing. This allows for an alternative to pure representationalism or pure dispositionalism: believing is representational, whereas beliefs are a function of this activity in real and counterfactual circumstances. Occupying this middle ground preserves something important about representationalism, namely the idea that belief happens in the mind and can be real even if it has no impact on action. At the same time, the planning account sides with the dispositionalist in allowing a looser relationship between thought and the total set of believed propositions.

The structure is as follows. In §2, I bring out two features of non-occurrent beliefs, and in §3, argue that these cause problems for representationalist accounts of belief. These problems

---

1. This is a penultimate draft of a paper to appear in *Philosophical Perspectives*. 
indicate a desideratum for accounts of belief: to be psychologically grounded but not dependent on representational form. §4 proposes a simplified model of planning, which forms the basis for an analogy between belief and planning in §5. In the final section, I discuss some upshots and challenges for the proposal.

2. On the intuitive view suggested by English idioms, believing means being in a state of having a thing. Contemporary analytic philosophers tend to adopt a modified version of this idea: belief is a combination of an attitude, something like taking-to-be-true, held toward a proposition. This thesis might be more or less representationalist, where pure representationalism locates thinks of belief as a representational state with propositional content and a connect, psychologically real, attitude, and the alternative extreme, dispositionalism, instead understands belief as a tendency to act and think in certain characteristic ways.

While the debate between representationalism and dispositionalism is fraught and ongoing (see Quilty-Dunn & Mandelbaum, 2018; Schwitzgebel, 2002), in what follows I will only argue against representationalism, since it will serve to highlight a challenge that provides the inspiration for my positive picture. Another reason to start with representationalism is the thought that it would in some sense be nice to be representationalists, if we could: that is, if there was a reasonable theory of belief that located beliefs in real, psychological states, we’d have a weak reason to prefer such a theory over an equally reasonable dispositionalist alternative just because representationalism involves a unification between psychology and philosophy, and provides a deeper explanatory footing for our theory by grounding beliefs in real states of the world rather than modal dispositions over mental states.

I’ll start by introducing two relevant dimensions of non-occurrent beliefs, assuming a representationalist view for reductio. First, belief, as well as representational states such as preference, supposition, and so on, can be arranged on a continuum of representational stability. Let’s say that Clelia believes the post office is closed on Saturdays. She formed this belief a long time ago, and perhaps not even at a precise moment of judgment, and tomorrow she will call it to mind when we discuss sending a gift to her sister. But how should we describe her state of mind today, when she’s not thinking about the post office at all? A view that ascribes high representational stability would say that she has, even today, a representation of the post office being closed, and that this very same representation will be used tomorrow when she does start thinking about the post office. This representation is stable in the sense of remaining in the same condition over time.
On the other end of the spectrum, we might say that today, Clelia doesn’t have a representation at all – tomorrow, when she starts thinking about sending the present, she will build up her belief from scratch, taking into account the context, her memories, and other background information. In this case, her belief is fully ad-hoc and occupies the opposite end of the spectrum from a highly stable, persistent belief. Of course, Clelia’s belief likely does not just pop into her head from nowhere – it is surely built on information carried forward in time from her past, as well as current information. But it is a new construction in the sense that these other information sources are not themselves beliefs, a substantial portion of the information is contributed at the time rather than in the past, and none of them plays the role of a single belief that the post office is closed on Saturday that travels through time, even when she’s not thinking about it. In our terms, this belief is representationally unstable.

Representational stability picks out one dimension of occurrent and non-occurrent belief. I’ll contrast this with a second, partially orthogonal dimension. Let’s say Clelia’s non-occurrent belief, as many beliefs would seem to be, is moderately representationally stable. First, Clelia might today have a sort of outline of a belief, an impoverished version that is still the same kind of state: a representation of a way the world could be. Then, tomorrow, she would call to mind this outline and fill it in based on the context and other information sources. Thus Clelia today might really believe the post office is closed on Saturday, but when she thinks about it tomorrow, we would only then be able to answer more precise questions: does she believe it of the US postal service only, or of others? Is her belief about the building she is pointing to, or about the whole institution? And so on. I’ll call this the “sketch” scenario, since her belief maintains the same representational form, but is somehow coarser, impoverished, or less vivid that the original, occurrent belief, not unlike the relationship between a complete painting and a sketch in that both depict the same scene with the same major objects, but differ in their degree of detail.

But this is not the only way a non-occurrent belief might differ from its occurrent counterpart. We might just as well imagine that Clelia today believes in virtue of a more diverse category of mental states and activities: for instance, rather than an outline, she might represent pieces, so to speak, of the full belief: an image of a sign that says closed, or any other component of the representation that does not possess the full form in the way an outline does. To put it another way, she might have thoughts that connect over time, that will become part of the belief tomorrow and were part of the belief in the past, but that do not represent the proposition in question.
This “pieces” scenario can in fact take the form of propositions. For instance, if she were to represent that ‘the post office has the same hours as the bank’ and ‘the bank’s hours are such-and-such’, these would not be impoverished sketches of the occurrent belief but instead pieces from which it would be reconstructed since they differ in structure to the complete belief. Assembling a belief from its “pieces”, as will become relevant later, has something in common with inference.

We might view perfect representational stability as a special case of the “sketch” scenario – one in which the sketch is perfectly details. But pieces and sketches do not always differ in terms of stability – both ways of believing can encompass a wide range between, on the one hand, representations that remain meaningfully similar in content over time with minimal variations based on context, and on the other, representations almost as thin as a totally ad hoc belief with no internal connecting thread between instances. Rather, they differ in what we might call representational form.

In summary. On the traditional view, where beliefs are objects that are held, we can locate theories of belief along a spectrum of continuity across time. Theories with high representational stability hold that to have a belief over time implies a relationship to a representation such that this representation is present over time between instances of the belief being made occurrent. Theories with low representational stability hold that to have a belief over time implies no such thing, and is usually realized by a mental process that reproduces the belief anew when made occurrent, drawing on a mixture of past and present information sources. We then distinguished a distinct dimension, representational form, and two ways beliefs might change in representational form when non-occurent: believing between occurrent episodes involves an outline of the belief that is impoverished or vague, or a set of pieces or fragments which do not themselves share a propositional, representational form with the original and reconstructed beliefs.

So far, I’ve entertained a conceptual question: what would it take to have a belief? But of course there is a highly related empirical question: in which of these ways do we tend to hold on to representations over time?

While a full answer to this question is outside the scope of the current paper, I’ll note several cases in which psychologists have posited a pieces view, with the aim of convincing the reader that this a serious, live possibility for many domains of belief.

First, in the literature on concepts, some have argued for exemplars (alone, or as a component of prototype theories, e.g. Murphy (2016)), on which a concept has a structure of a few key
instantiations, and new items are classified by their closeness to those exemplars. On this view, the concept that allows me to make judgments such as “tea is a drink” is itself not a set of propositions, but a set of paradigm instances of “tea” and “drink” (along with a more complex prototype structure, on some accounts). Thus these propositions would seem to be stored, at least sometimes, as latent derivations of the concept rather than as sketches or complete propositional representations.

Second, in the case of intuitive physics, many psychologists believe we make judgments through mental simulation. For example, Gerstenberg & Tenenbaum (2017) explore the way people make judgments about ordinary physical issues, such as where the two balls will go after their collision (figure 1).

On their model, while I may have seen or thought of this particular kind of collision before, I tend to answer questions about it not by bringing up a library of past knowledge or experience of previous collisions, but by running a mental simulation. A simulation is a representation of an imagined event unfolding in time, in this case something like a mental movie of the collision. This simulation comes from a model of the world that encodes basic physical principles but that I access through simulation. Thus the principle stored representation is the model itself, and when

![Figure 1: To predict the resulting ball positions when the plain ball collides with the striped ball at the angle shown, it’s natural to use mental simulation.](image-url)
I extract answers in specific cases, the information is transformed to a new representational form: a prediction about a specific instance.

Other similar cases include the use of mental models in theory of mind (Jara-Ettinger et al, 2016) and the use of sampling processes in causal judgment (Icard & Knobe, 2016). In each case, I don’t mean to suggest that psychology has definitively proven that a pieces-like representational form characterizes non-occurrent belief, but instead that across a wide range of domains, it’s plausible that such a form is widespread. This is unsurprising, because shifts in representational form affect the efficiency of storage and retrieval, as well as the distribution of errors, and so we should expect variations in cognitive resources and risk to make different representational forms optimal in different circumstances.

Thus both pieces and sketches are possible realizations of non-occurrent belief that can have the same degree of representational stability. There is considerable evidence that the pieces scenario is a better description of some of our beliefs in domains such as intuitive physics, theory of mind, and decision-making.

3. The plausibility of variation in representational form is the basis of a dilemma for a representationalist theory. Following the previous examples, I assume that many of what we normally call our beliefs do not have full representational stability. Of these, some but not all persist across time in the representational form of sketches: vague or coarse-grained versions of the occurrent belief. Instead, many take a different representational form: that of pieces of the belief that do not themselves express a version of the relevant proposition but are connected with other pieces and the context when a question is called to mind.

Now let’s stipulate that before she asks herself the question about the post office, Clelia has (in the traditional sense of metaphorically possessing) some pieces of her belief, perhaps an image of a sign, an incomplete representation of the hours of operation, or similar. At that point in time, in virtue of what does Clelia count as holding the belief?

One possibility is that she does not actually count as having the belief, when the pieces are non-occurrent. This is an undesirable answer because we think of people as contradicting themselves, in situations where they, for instance, only think about how nice it would be in Florida while not bringing to mind their background knowledge that there is a hurricane in Florida right now. This kind of contradiction, at least if we interpret it straightforwardly, points to an incoherence between an occurrent and non-occurrent belief. Were we to classify the pieces scenario as not meeting the standard of belief, this type of self-contradiction would be no different from other
cases in which we have consistent beliefs, but sometimes do not draw on all available information: for instance, I might be making a plan to fix my bicycle tire without taking the time to figure out which route to the repair shop would use the least gas. Incoherence between an occurrent belief and the pieces of that belief seems different. Here, the non-occurrent pieces are not just inferentially related to a potential belief, but are the connecting thread, in non-incoherent cases, between my actual beliefs over time.

More generally, since occurrent thoughts represent a small fraction of the time and influence of a doxastic state, a too-strenuous standard of what counts as a belief will rule out many relationships between mental states, such as (in)consistency over time, longer-term reasoning, and other transitions that go beyond just occurrent thought.

Could the answer be that she counts as holding the belief now in virtue of being disposed to assemble it in the future? Even though we are now assuming representationalism, we might allow for a more dispositionalist answer to the question about non-occurrent beliefs. But suppose we take this route: only occurrent beliefs are held in virtue of representations, beliefs like Clelia’s now are instead held in virtue of being disposed to hold representations at the appropriate time (for instance, when asked). This answer limits the scope of the traditional view, such that either only beliefs held occurrently are really “held”, or instead, we should weaken our notion of holding a belief to capture both forms of relationship, the dispositional and the representational. The former answer goes against our assumption that many everyday beliefs are of incomplete representational form, since we would have to either allow that these everyday states are not beliefs, or that they are not of incomplete representational form. The latter takes us a bit afield from the traditional view proper, and raises some challenging issues.

How could a single concept of belief be satisfied by both representational and dispositional instantiations? Right away, we should rule out the possibility that this works by merely weakening the theory of belief ascription to a dispositional one: this both goes against our initial assumption of moderate representationalism, and also would imply that representational ascription is strictly stronger than dispositional ascription. To put it another way, it is not likely that anytime a state S is classified as belief under a representational criterion, that same state would always be classified as such under a dispositional criterion. For instance, I might hold a representation that animal souls possess a deep and mystical unity, and thereby count as believing it (provided I have the relevant relation to that representation). And yet, this belief may be so imprecise, hard to articulate or disconnected from other thoughts and behaviors that I would not be disposed to act, think, or even speak differently were I not to hold such a representation. More
generally, on many accounts of how these two types of ascription (representational and dispositional) function, ascriptions depend on quite distinct properties and therefore it would be surprising if the dispositional ascription were to be always applicable whenever the representational ascription is.

But what about some other way of linking the two types? We might embrace a disjunctive theory of belief ascription, or similarly, a family-resemblance or cluster-based theory on which both representational and dispositional characteristics are sufficient for believing without there being a more abstract notion of belief under which both fall. I won’t argue against these views here, but only suggest that they seem somewhat objectionably ad hoc, and fail to answer a question that would be at least desirable, if not necessary, to answer: what is held in common between Clelia’s belief now and her belief tomorrow?

This first way of responding to the question gave up on making a (representational) possession out of the pieces of a belief. The second response instead attempts to do so. What do these pieces have in common with the belief she’ll form out of them? Some options include: (a) they will be causally related to her future state, (b) they will be constitutively related to her future state, or (c) they already relate her in the right way to the proposition, regardless of what will happen in the future. Of course, there may be a cluster or combination of these connections that is relevant or necessary for her to believe now.

Options (a) and (b) both rely on what will happen in the future. This generates a predictable problem: imagine we have two identical Clelias right now, but only one of them will happen to be asked about the post office, and thereby count as having the belief now. This picture looks most appealing when we are discussing a stretch of time in the past, and asking, what did Clelia believe between then and now? From that perspective, the contingency is removed, and the explanatory, causal, or constitutive relevance of the past state is foregrounded. Nothing about these views contradicts the basic idea of the standard view, as it is equally possible to give a theory of, for instance, being the owner of a property, that takes the relevance of the paper deed to depend on the future contingency of whether it will continue to be honored.

Option (c) raises our initial question, in virtue of what are the “pieces” of the belief related to the belief, in a new form. Setting aside the answers above, that a current disposition of the whole person or future link does the job, we now need an answer that draws on only the current status of the “pieces”. That is, they should stand in a relation to the full belief that makes sense of why we take Clelia to hold the full belief now, but without dependence on any future contingencies.
So what might this connecting thread be? We have the intuition that a function from pieces to complete belief will be radically under-determined - surely there are many beliefs that could be formed from the “pieces”, especially given the role of the person’s unique cognitive dispositions and inclinations. Instead, I’ll suggest an alternative. The traditional view is forced to explain how the belief could be implicitly present in its pieces, such that when you put the pieces together, the implicit belief is actualized. On the contrary, I hold that the way of putting together the pieces just is the belief. More precisely, it is the believing: an activity that underlies belief ascriptions.

Someone might object that these negative arguments take the metaphor of having the belief too seriously. There is of course a spectrum of representationalism, and it would be a mistake to take the most extreme version as our only target. Here I’ve argued against only forms of representationalism that are sensitive to representational form, and therefore care about the difference between sketches and pieces. We can contrast:

**weak representationalism:** holding a belief that P entails having an overall representational state or some subset thereof that maps onto the content P

with:

**strong representationalism:** holding a belief that P entails having a representation with the content P*, where P* is a suitably related proposition or P itself.

The arguments above do not tell against weak representationalism, since the key feature of the non-occurrent pieces scenario is just that there is no particular representation that corresponds to her belief, nor a suitable function from those particular “pieces” to the content of the belief. That is not to say that there is no suitable function from her complete state of mind and dispositions to the content of the belief. But while strong representationalism might seem implausibly literal, it is defended by some (e.g., Mandelbaum forthcoming, 2019; Mandelbaum & Bendaña, 2020). More importantly, attending to this problem with strong representationalism will set a key aim for an alternative: to explain how pieces of non-occurrent belief relate to occurrent belief, and to allow for belief to stretch over time and across shifts in representational form.

4. The spectrum of representational form raised a challenge for strong representationalism. In this section, I’ll begin building an alternative with a basic model of planning.

Consider what it means to plan to go to the store. Typically, that you will definitely end up going to the store is not enough to count as planning to go, nor is it sufficient that you will definitely go
and you know it. A different kind of cognition is involved than mere awareness. And yet, planning to go to the store does not require having a mental object, the plan to go to the store, nor is having that object sufficient to count as planning. To see this, imagine someone who has just turned to the question of what to do this afternoon, and wonders whether she should go to Food City or El Super. She has not yet created a plan -- she’s only asked the question. And yet, intuitively, she is planning to go get groceries, even though her plan is barely begun. In fact, one might argue that the process of planning is usually over once a plan has been produced (though equally, we might call certain unfinished plans that characterize much of planning real plans, in which case, most of planning would involve having a plan). Likewise, if someone has a mental representation containing all the correct internal features of a plan to go to the store, that is not enough to count as planning: she might be considering what someone else might do, or entertaining a counterfactual, and so on.

Thus, planning is an activity, to count as planning requires a certain kind of thought (not merely action or disposition to think, act, or feel), and planning does not reduce to having a mental representation, a plan. I will import these features to the case of belief, in order to build an alternative to the traditional view. But first - what kinds of thinking are characteristic of planning?

To answer this, let’s look at one model of planning, drawn from Ho et al (2022). Imagine an agent who is planning how to climb a mountain. Her goal is to reach the mountain, but this does not exhaust the way she thinks of the problem: she also faces a task, which Ho et al. understand as an enumeration of possible acts and states of the world and their associated transitions. She also has a task construal, which is how she represents the task to herself: rather than thinking of all possibilities and all transitions, she will instead represent a smaller subset of options along with sub-goals and estimates of outcomes. For instance, she might break up the task into finding gear, packing, and climbing, and within each sub-goal think through and weigh a few options. In planning, we might imagine all she does is evaluate the options specified in the task construal and select those she deems best. However, Ho et al. provide a model (along with some empirical confirmation in a grid-world domain with human subjects) that allows the task construal to be updated during planning as well: thus planning is not just a top-down elaboration and evaluation, but a process of weighing both which thing is best, as well as how to set up the problem you’re trying to solve.

Drawing on their approach, we can build a basic model of planning as show in the following table:
Planning

Goal (or utility function) – an assignment of value to outcomes or outcome-types

Task - full specification of possible acts, states, rewards and a transition function

Task Construal - an updateable and simplified task representation

Planning - progress through task construal and evaluation

Plan - the partially specified output of planning at a given point in time

On my adaptation of this model, planning to reach the mountaintop often involves only representations asking and answering questions such as “which backpack do I need?” and “what route to the summit is most efficient?”. This in principle does not require an explicit representation of the goal, to reach the mountaintop. More specifically, take any goal state G. As has been observed everywhere from explicit planning (Bratman, 1999) to motor control (Wolpert, 1997), this abstract state is typically broken into a hierarchy of sub-goals $i_G\ldots n_G$, with their own corresponding sub-goals. Within this hierarchical structure, an agent might consider $i_G$ explicitly without explicitly representing it as a sub-goal of G. Often, we say $i_G$ is a sub-goal just in case it is embedded in a system that treats it as a sub-goal, and likewise for tasks, task-construals, and plans. In fact, explicit planning seems to rarely involve an explicit representation of the very highest level of the hierarchy – after all, that level is not where the action is usually happening.

However, we should resist the urge to see all planning as descending in abstraction. Thinkers may encounter troubles at the $i_G$ level that cause them to revisit and revise G, at which point G or at least questions and concepts at the level of G, would typically be explicitly represented. Planning on this model has in effect four directions. Within a task construal, planning can flow from abstract plans to concrete specifications and from concrete specifications to the revision of more abstract plans. Two other directions are also possible: from a task construal to the beginning of
This model has several key features. First, planning is an activity, and one that is not reducible to having a plan. In fact, having a plan is neither necessary nor sufficient for planning: in the former case, I might have fallen into total uncertainty through the process of planning, and in the latter, I might have been handed a plan by another person without thinking of it myself. Second, planning is cognitive: it is not a way of acting, but a characteristic, and informationally specified, series of computations.

Third, the plan itself is an abstraction characterizing where I am in the process, not itself an artifact. This is particularly clear in cases where I am interrupted in the middle of planning before updating my explicit plan representation: in these cases, the model would identify my plan with, loosely, whatever I would commit to given my current progress, rather than whatever is already explicitly represented. This specification is by its very nature incomplete and open to different interpretations (for instance, what I actually would commit to, vs. what I should commit to).

On my account, planning is primary and plans derivative from the activity of planning. This may strike the reader as counter-intuitive. I starting by arguing that planning doesn’t reduce to having plans. But even if this is true, we might still ask: doesn’t planning always end in a plan? I think this idea is better captured by saying planning aims to end in a plan. The framework I’ve presented explains this normative connection. When I am not planning under pressure, in the idealized situation where I am planning only before acting rather than during it, and I come to a stable, settled idea of what I am going to do, the idealized plan (“what I would say if you stopped me right now”) will match this settled idea. There will be a single representation that is my plan both in the sense of an internal map of what I am going to do, and in the sense of an abstraction that summarizes my place in the process of planning. So the claim that planning aims to end in a plan is true when interpreted as saying that planning under ideal conditions ends in a plan. On this interpretation, however, it does not imply that any real case of planning actually ends in a plan, nor that in any actual situation, planning should end in a plan.

5. This model of planning will now form the the backbone of a proposal for belief and believing. Instead of the goal of climbing the mountain, now imagine that our agent is thinking about birds and how they construct nests. In the belief domain, we may not need goals per se, and especially

2 Assuming I am the type of creature who represents cognitive commitments of this kind explicitly.
not deliberately pursued, voluntarily adopted ones, but we can think of an agent’s learning as oriented toward an aim. Perhaps these aims are always subsets of the aim of believing the truth, but it seems like in ordinary life, we are oriented toward figuring out the truth with respect to some particular topics at a particular time, rather than merely generically oriented toward truth.

The counterpart of a task, for belief, is a learning problem: an enumeration of possible observations, associated states of the world, and transitions between them. Unlike in the case of action, inferences will rarely feed forward into (relevant) changes in states of the world. In principle, we might also include epistemic acts such as experimentation, which do of course determine future evidence, though this threatens a full collapse of the belief problem into the planning problem. Further, trying to figure out what is true characterizes a certain kind of thinking, a kind of thinking distinct from trying to figure out what to do to figure out what is true. For this reason, I treat the learning problem as a matter of figuring out how the world works, or a more narrow version thereof.

A task construal corresponds to a query: a specification of sub problems or rather sub questions, features calling out for explanation, or elements of the broader aim that need to be understood. Just like a task construal, a query is the agent’s own simplified way of understanding the problem, and following Ho et al’s model, the query can and does shift during the process of believing.

<table>
<thead>
<tr>
<th>Planning</th>
<th>Believing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goal - (or utility function) – an assignment of value to outcomes or outcome-types</td>
<td>Goal – (or utility function) – an assignment of value to epistemic outcomes or outcome-types</td>
</tr>
<tr>
<td>Task - full specification of possible acts, states, rewards and a transition function</td>
<td>Task – full specification of possible observations, states of the environment, truth values and a transition function. A learning problem</td>
</tr>
<tr>
<td>Task Construal - an updateable and simplified task representation</td>
<td>Task Construal - an updateable and simplified question representation. A query</td>
</tr>
<tr>
<td>Planning - progress through task construal and evaluation</td>
<td>Believing - progress through reasoning and query evaluation</td>
</tr>
</tbody>
</table>
Plan - the partially specified output of planning at a given point in time  
Belief – the partially specified output of reasoning at a given point in time

Thus on this model, an agent counts as believing that birds create nests in virtue of a certain kind of thinking: thinking that formulates and reformulates queries and evaluates the possible answers to these queries, as an activity organized and oriented toward that aim.

Note that in general, the activity of believing that P will not be identified with the activity that aims to determine whether P. Instead, just as we plan to go to the store by trying to figure out whether there will be traffic, we believe that bird create nests by trying to figure out the types of nest, by explaining why a hummingbird uses a certain nest material, and so on.

As I’ve noted, asking oneself about sub-goals or sub-plans is the kind of thinking that constitutes planning. Planning has long been discussed as having a hierarchical structure, and it seems plausible that all planning, or paradigmatic planning, follows this structure. However, the same cannot be said for belief: some beliefs can be further and further specified (birds create nests, hummingbirds create dainty nests, and so on) but others do not obviously lend themselves to this treatment. For instance, consider my belief that triangles have three sides: this belief is not a coarse-grained version of a more specific belief, except in the trivial sense of ruling out fewer possibilities than, say, triangles have three sides and Tucson is smaller than Phoenix. However, my belief that triangles have three sides can form the basis for a wide range of downstream predictions, questions, and explanations and so in this sense it can be captured by our account.

In general, an activity of thinking T will count as a way of believing that P in proportion to the degree to which T reflects a commitment to P. This activity of thinking will not just include reasoning, but also explaining, predicting, analyzing, questioning, and so on -- everything that goes into building and using models of the world, including evaluating the efficacy of the model, use, and task construal as we go along. This relationship between thinking and believing might sound circular or uninformative, but two features of the planning model make the connection more specific.

First, starting with the scheme above, we can examine a thinker’s mental activity to see which doxastic goal and construal best reflects her thinking. Once we have these specified, we can use the schema to infer belief contents. These will be a set of many contents, with some being a close
fit by which we mean indispensable to describing the construal. In this category we’ll find things like immediate presuppositions of the questions she asks herself, as her believing that cats are a kind of mammal is instantiated by attempt to explain why cats are the only mammals that show such a brazen drive to curiosity. In a hierarchically structured task, these close contents will include higher-up generalizations of the questions she’s considering, for instance believing that cats exhibit unique behaviors. For these contents, we won’t have to look very far to determine what she is believing. As we examine more distal contents, there might be some vagueness. Thus the link between the activity of believing and the contents of belief is in many cases a direct and obvious logical question that attaches a content to a way of thinking, though we can also distinguish a gradation in closeness such that other contents are more loosely and perhaps vaguely ascribed.

Second, in many cases (and on some views of believing, all cases), the believer herself will be self-conscious of her project. This means that she understands herself to be believing under a certain description, rather than the theorist merely ascribing a belief to her on the basis of implicit thought. At least when she has a good understanding of herself, this gives us extra resources to make the activity of believing a specific content concrete. In many cases, then, a thinker will be explicitly moving through the belief task construal, towards an explicit goal of understanding a particular domain, and under the auspices of an explicit set of commitments. On our view, this case has the key elements in common with others that involve more implicit thought, though the explicit case is more determinate with respect to content. However, the core of our theory is compatible at least in principle with the requirement that believing be explicit. Critically, believing that P can then be explicit without the content that P being explicit.

That said, many actions have nothing to do with plans: they are immediately triggered by the environment, and neither thought-out nor linked to a more complex ongoing activity. Accordingly a counter-intuitive result of this view is that “beliefs” that consist in directly forming a thought based on the environment and then never thinking in ways that reflect a commitment to that content again, will not count as beliefs at all. For instance if I see a turtle, think “that is a turtle!”, and never go on to consider the event (or closely related events/concepts) again, including never forming a memory of the event.

I’ve argued that the activity of a certain kind of thinking is psychologically fundamental to the state of belief. This activity of thinking, believing, involves computations over psychologically real representations. But these representations are not themselves beliefs, or rather, they need not be. Why wouldn’t we want to call these representations beliefs? I’ve already given one answer: in
many cases, we’re believing already before the belief is encoded in a representation. But conversely, we think with a variety of representations to which our commitment falls short of belief. For instance, I use a cognitive map of my environment while understanding that as a provisional simplification, it’s simply false. I compute the consequences of suppositions. I consider memories whose contents are contrary to what I really believe (such as in Mazzoni et al 2010) or which I hold at arms’ length and consider. My point is not just that we adopt many different attitudes toward occurrent representations. Instead, the things we compute over are simply representations, rather than complexes of attitudes and representations. The attitudes come in when we try to understand why I am computing in this way, and this question has a holistic answer rather than one that decomposes into discrete attitudes held toward particular representations.

In summary. Just like the case of planning, our theory of belief holds that believing is an activity not reducible to beliefs. Believing is a computational and presumably representational cognitive process. Belief, on the other hand, is an abstraction over that process which is only partially specified.

6. In this section, I’ll look at some consequences of the view, and a few objections.

Let’s return to Clelia’s case, and the question of what makes a non-occurrent state a belief. On our view, there is no difference between the “pieces” and “sketches” views -- both can be equally relevant to the activity of believing, since we make no special place for a representation whose form (and content) matches the belief. We think this is the right result, since form is orthogonal to stability, and stability is more apt to capture the use of belief. But this raises two further issues.

First, one difference between the “pieces” and “sketches” scenarios might be that the former involves a kind of inference, whereas the latter merely a filling in. If so, by marking no difference between the two scenarios, we might be accused of eliding inference and memory. To this, our response is that many models of memory already include inference-like features (e.g. De Brigard (2014), Aronowitz (2019)). Perhaps softening this distinction is actually a positive move towards a more complete picture of memory.

Second, the question of occurrent vs. non-occurrent does not immediately apply to the belief-as-planning view. When a thought is occurrent, it is before the mind’s eye, but activities cannot be “before the mind’s eye” since they are ongoing complexes of extended processing. Thus activities are instead things we are more or less actively engaged in. For instance, I am very actively engaged in training for the marathon when I am at a practice run, somewhat less when I am idly
searching the internet for new shoes, and not at all when I am sleeping. This, as opposed to the
occurrent/non-occurrent issue, is a question of degree. So we are already in quite a different
position with respect to what was non-occurrent belief and is now inactive or less active belief.

By the same token, a belief will be active much more often on the planning view than it would be
occurrent on the traditional state view. This is because not only is being active a question of
degree, but so is the relevance (and hence ascription) of the belief itself to my current cognitive
activity. This forms a marked departure from the standard view, on which only bringing to mind
that particular belief can make some content occurrent.

Putting all this together, our view evades the problem of non-occurrent belief, and predicts no
difference between cases of “sketches” and “pieces” of belief that have the same degree of
representational stability. The view, however, still marks out a difference of degree in actively
believing. This is not a complete solution to the problem, since similar questions can be raised
about what is going on with non-active believing, and how to relate it rationally to active
believing, but we are at least now debating a smaller number of cases.

7. In criticizing a different sort of representationalism, Gilbert Ryle wrote:

In unconscious reliance upon this dogma theorists and laymen alike constantly construe
the adjectives by which we characterise performances as ingenious, wise, methodical,
careful, witty, etc. as signalising the occurrence in someone’s hidden stream of
consciousness of special processes functioning as ghostly harbingers or more specifically
as occult causes of the performances so characterised.

I have argued the same error is made in the case of belief. The representationalist is looking for a
further cause, a representation that P, that unites and relates the various processing, bits of
information, and other parts of cognition whereby we take the world to be a certain way. The
planning theory of belief is a fairly representationalist, fully cognitive theory. But rather than
belief being a state that lies behind thoughts, causing them to work in certain ways, I have posited
instead that belief is an abstraction over more fundamental believing processes. Believing, this
more fundamental notion, is a kind of thinking that reflects, but need not encode, a commitment
to a content. We can preserve the representationalist idea that what’s going on right now in our
minds is the core of believing while giving up on belief as a “ghostly harbinger” or “occult
cause” of our other thoughts, of the activity of asking ourselves what we should do, why things
happen, and what will happen next.
Acknowledgments. This paper emerged from joint work with Tania Lombrozo. I’m also indebted to Reza Hadisi, Harvey Lederman, Jennifer Nagel and Julia Staffel for helpful comments.

Bibliography.


