

This is a penultimate draft of a chapter forthcoming in Schwitzgebel & Jong (eds.) *The Nature of Belief*.

## A Planning Theory of Incoherence in Belief

Sara Aronowitz

### Introduction

In the psychology of religion, researchers have documented a series of cases across cultures where adults as well as children seem to operate with two clashing frameworks: for instance, the belief that witchcraft causes disease, and the belief that germs cause disease (Legare & Gelman, 2008). Susan Gelman puts the problem of explanatory coexistence in this domain as: “Why do some beliefs that contradict science persist and others disappear?” This question can be asked even more generally: why do some beliefs that contradict our dominant explanation of the world persist and others disappear? At least in some domains, psychologists have argued that the coexistence of seemingly contradictory beliefs is not an initial error that takes a long time to correct, but instead reflects a sustained process of supporting both sets of beliefs. In other words, these contradictory beliefs are not just held but maintained.

The question raised by Gelman is an open area of active research. It has an empirical aspect (which contradictory beliefs are actually maintained?) as well as an epistemological one (what are the learning costs and benefits of maintaining contradictory beliefs?). However, two prominent theories of belief, pure dispositionalism and pure representationalist, do not help us answer the question -- and in fact even encounter difficulties in *formulating* it. The aim of this chapter is to show how an alternative conception, the planning theory of belief (Aronowitz, in press), does better.

The planning theory is built around an analogy between belief and planning. Incoherence in planning is widespread and not particularly mysterious. I start by considering cases where a person is actively planning two different sets of plans that could not (or would not) be enacted together. For instance, you might be planning a hike over the weekend, and at the same time planning a lazy day at home. In some cases, you have a settled understanding of how the plans fit together: if it rains, you’ll stay home, and if it’s sunny, you’ll hike. However, at least sometimes, we plan in this overlapping way without first determining how the plans fit together. In this case, it seems intuitive that we do plan in jointly incoherent ways, sometimes at the very same time, but that there are clearly cases where planning is more streamlined and coherent. That is, the analog of Gelman’s question - why do some plans that contradict our dominant plans persist and others disappear? - makes perfect sense.

My suggestion will be that several features of the structure of plans and planning ground our understanding of contradictory (or “double”) planning. First, the activity of planning is a process of assembling plans but does not require having an artifact, “the plan”, held in one’s mind. On the other hand, planning is not merely a disposition to act a certain way, but does normally

involve a distinctive and real cognitive process. If we import these two features as part of a theory of belief, I argue, the foundation can be laid for a fruitful answer to Gelman's question.

The structure is as follows. I'll first present the problem of explanatory coexistence and connect it to belief. In Section 2, I sketch the difficulties faced by dispositionalism and representationalism in dealing with incoherence, and discuss the relationship between explanatory coexistence and fragmentation. Section 3 explores incoherence in planning, and Section 4 shows how the planning model allows a better understanding of incoherence. I conclude by considering how the foregoing contrasts with a long philosophical tradition of under-appreciating incoherence.

### **1. Explanatory Coexistence**

Explanatory coexistence is a phenomenon in which people explain the world through incompatible frameworks. At the level of this operational definition, explanatory coexistence is a behavior, but we may also stipulate that this explanatory behavior is sincere in that the person is not explaining only for the benefit of others, or otherwise overtly saying something they don't believe. However, we should leave open the possibility that these explanatory behaviors reflect acceptance, rather than belief, or some other attitude that falls short of endorsing the explanation as literally true<sup>1</sup>. It also leaves open the possibility that the two frameworks in question might not be identified by the person herself as incompatible, either because she has not fully considered the issue or because for her they are subjectively compatible, for instance intelligent design and evolution might be reconciled through a personal theory that evolution itself was designed by God. This way of defining the phenomenon is important because it will give us a more neutral starting point from which to ask what this behavior says about belief.

I'll now survey a small portion of the work in this area – see Sommer et al (2022) for a broader look at evidence of incoherence in belief generally. Astuti (2002), uses both cognitive psychological experiments and qualitative fieldwork to describe two different frameworks for the understanding of mind and body in her participants, Vezo adults and children in Madagascar. She finds that adults produce two seemingly contradictory behaviors: first, they explain physical and mental traits of children by referencing the proximity of adults who are often not biologically related to the children (e.g. “she must be tall because her family friend is tall”). Second, when questioned about the inheritance of traits from biological and adoptive parents, they categorize physical traits as biologically inherited and mental ones as acquired from adoptive parents. In the adults, then, Astuti sees a sophisticated pattern of a social custom of treating parenthood as always acquired overlaid on top of a “dualistic” understanding of inheritance. In children, the results are more mixed: the majority of the children in her study did not clearly show the body-biological/mental-acquired pattern, and when asked about a case of adoption between a familiar context and an unfamiliar one, they did not consistently generalize from one to the other. In this set of studies, we see that incoherence obtains in adulthood and does not seem to arise merely from one framework (presumably, dualism) being innate and a later framework (social acquisition of all traits) learned on top of it. Instead, she suggests that at some point between early childhood and adulthood, the participants learned both a dualistic pattern of classification and the social-acquisition model, alongside an understanding of when to use each. This seems

---

<sup>1</sup>See Van Fraassen (1980) for a classic discussion of the option here that fall short of full belief in a theory.

like the best explanation for why the adult participants were fluent in both ways of explaining descent, whereas the children were competent at neither and relied more on exemplars rather than generalization.

A second example is the explanatory coexistence between scientific and folk-scientific beliefs. Shtulman & Legare (2019) looked at scientific beliefs in several domains in American college students where scientific categorization departs from intuitive characterization. For instance, rivers might be intuitively categorized as alive while scientifically categorized as not living, and mold might be scientifically categorized as alive while intuitively categorized as not living. In these conflicting cases, their participants are slower to verify scientific statements and make more errors. They propose two models to explain these findings. The low-level associative model predicts that the categorization of rivers as living persists in associations, especially those present in the lexicon of English itself rather than fully internalized in speakers. The high-level theoretical model instead holds that intuitive theories are really theories, that is cognitive structures that have a logic and inner workings that support generalization and inference<sup>2</sup>.

To see the difference, consider the statements “fire is alive” and “oaks are alive”. Fire, according to their corpus analysis, is more likely to be referred to as alive than oaks, which makes sense given the use of metaphorical language around generating, killing, and resurrecting fires. According to the low-level account, fire would then be associated with life. But fire of course is only metaphorically alive; according to a sort of folk theory of life, which Shtulman & Legare take to involve moving on its own, sensing the environment, and having goals, oaks are rated as more alive than fire on all of these dimensions but especially in sensing their surroundings. Thus according to the high-level model, oaks would be more closely associated with life than fire. Shtulman & Legare ultimately find evidence that slightly favors the high-level account, but I put forward their results in part to open up a question: once we observe a conflict in explanatory behavior, is this best explained by two explanatory frameworks really coexisting in the mind in the same way, or might one have a sort of lexical or associative low-level lexical existence? Further support for the high-level case comes from findings that suggest young children do not have any better grasp on the non-scientific option, as we saw in Astuti’s study, and in fact sometimes take longer to achieve competence at the non-scientific framework (Legare & Gelman, 2008).

There is a lot more to be said about explanatory coexistence. For now, I want to draw out a few threads in this area of research. It seems that all of us experience some degree of coexistence: it is not isolated to children, non-Western societies, or those lacking in formal education. The attitudes we take toward the coexisting frameworks in some cases seem to be symmetric: for instance the Shtulman & Legare study and the findings in children from Astuti and Legare & Gelman indicate that we don’t merely have intuitive, non-scientific associations but likely have a true competing theory that is learned and richly structured. Third, while it is not at all obvious that we believe in either or both of the explanatory frameworks in these cases, both frameworks influence actions and thoughts in a way that goes beyond merely pretending as though the

---

<sup>2</sup>The relationship between intuitive theories and formal scientific theories raises a set of questions about incoherence in cognition vs science that I do not take up here.

frameworks are correct. And finally, while in many cases the conflicting theories have some separation into contexts, such as scientific beliefs belonging in school and dualistic inheritance beliefs belonging in private, this contextual separation is not absolute: Astuti notes that anthropologists have previously treated the socially-acquired beliefs of the Vezo as evidence that they are pure non-dualists across contexts due to the ubiquity of these explanations, and Shtulman & Legare find traces of non-scientific beliefs even when explicitly instructing subjects, who are college students in a psychology experiment, to verify the statements according to science.

This leads me to a set of desiderata for a philosophical account of how belief fits in to coexistence. Minimally, such an account should be consistent with the following:

**Incoherence.** the conflicting frameworks are not usually completely rendered consistent with one another but genuinely conflict both objectively and subjectively.

**Symmetry.** Explanatory coexistence is not always a matter of two frameworks in which we are asymmetrically invested, such as one believed framework vs one non-believed one.

**Partial contextuality.** Which explanatory framework is utilized is determined partly but not fully by context: at least some contexts trigger the use of both frameworks.

These desiderata indicate that explanatory coexistence is related to but not identical with what philosophers call fragmentation: the idea that a total belief state can be understood as consisting of subsets called fragments, such that some operations over these states are carried out relative to a fragment rather than the whole belief state. Fragmentation, unlike explanatory coexistence, assumes the attitude in question is belief. Both fragmentation and coexistence satisfy incoherence and symmetry. Fragmentation is often assumed to violate partial contextuality: fragments are fully contextual relative to an elicitation condition (Elga & Rayo, 2021) or question (Yalcin, 2021). Explanatory coexistence also refers to a behavior, the practice of explaining, which is both more specific than belief, since we have many beliefs that are not explanatory of anything, and more general, in that the attitudes we take toward our frameworks are not always belief or degrees thereof.

Explanatory coexistence clearly characterizes a lot of our mental lives: we employ ways of explaining that don't fully fit together all the time, and often we're aware of this. As we've seen, the next question is: why is this, and when does it happen? This is a psychological question but also a philosophical one, at least if there is a rational reason for coexistence. Given the centrality of belief and/or its close relative probabilistic expectation to theories of rationality, the natural question to ask is: how is explanatory coexistence connected to belief?

In the next section, I'll consider this question on two opposing understandings of belief.

## 2. A Challenge for Representationalism and Dispositionalism

In what sense are beliefs really encoded in the mind and/or brain? A pure representationalist claims that to believe that P is to have a representation that P to which you are related in the correct way (which you take to be true, or which you are committed to, etc). On the opposite end of the spectrum, a pure dispositionalist holds that you do need any particular representation, let alone one with the content P, in order to believe P: all you need is to be disposed to act in ways that reflect a commitment to P being the case. Most dispositionalists follow Gilbert Ryle in considering dispositions to act to include dispositions to mental or covert acts, those that have no clear outward sign such as doing a calculation in your head, suppressing an emotion, or imagining a scene. Some, such as Eric Schwitzgebel, go beyond the class of mental acts to include mental states, such as phenomenological seemings. Of course, many options exist beyond these two extremes, and other dimensions could be proposed to categorize theories of beliefs. For present purposes, I'll look just at a version of each extreme – not because I want to imply these are the most plausible candidates, but because the different sorts of difficulties they encounter with coexistence will reveal something interesting about what exactly is going on with coexistence.

First, the representationalist. It might seem that representationalism about belief would be quite suitable for understanding coexistence. I'll take Eric Mandelbaum and coauthors' version of representationalism as a model. On their view, what it means to hold a belief is to have an encoded representation that is used according to the psychological "rules of the game" of belief. For instance, beliefs are updated categorically whereas sub-doxastic states drift continuously, beliefs are connected to other beliefs in local "fragments", and we are attached to our beliefs in a distinctive way. When it comes to contradictory beliefs, the fragmented representationalist is committed to two claims:

**Fragmentation.** Rather than being stored in a single web, beliefs are stored in many, often contradictory, subsets of interconnected representations.

**Persistence.** We almost never cease to believe something, instead fragments are suppressed or become hard to access.

The functional role that this picture reserves for doxastic states suggests that many explanatory frameworks will be doxastic rather than subdoxastic for him: the sub-doxastic states, which are non-propositional and continuously updated are much like the low-level associations that Shtulman & Legare discuss, whereas the doxastic states would correspond to high-level theories. So in many cases, the fragmented representationalist would hold that we believe the tenets of both explanatory frameworks, just in different fragments.

Putting these claims together, of course, gets us the result that Gelman's question is meaningless: contradictions in beliefs do not ever really disappear. But there are several ways around this. First, the fragmented representationalist might argue that instead of eliminating our ability to pose Gelman's question, his account just pushes us to pose the question in a slightly different form. Namely, why are some contradictory fragments suppressed or made less accessible whereas others maintain their salience? The second response would be to argue that while

Mandelbaum and coauthors themselves might be committed to persistence, there is a better version of strong, fragmented representationalism that jettisons persistence and can set up an answer to the contradictory belief question.

The first approach could thus describe coexistence as incoherent fragments with a relatively high likelihood of activation, whereas a single unified framework would be a situation where all but one of the fragments has a very low likelihood of activation. This seems reasonable, because there is not much daylight between a framework that has been rejected and one that is in some sense present but is never retrieved. One way to ground this difference in features of the belief state would be to appeal to quantity: beliefs are retrieved more frequently when they are represented in more fragments.

Is quantity all that predicts when a fragment will be utilized? Another part of the explanation comes from the dynamics of retrieval:

The [sleeping-dogs] principle dictates that a fragment remains quiescent unless a) a search is triggered for its specific heading, and b) once that heading is located, searches cease. As long as inconsistent beliefs are housed in separate fragments, a sleeping-dogs principle dramatically decreases the likelihood of coactivating the inconsistent beliefs. (Bendaña & Mandelbaum, 2021)

This passage states that retrieving inconsistent beliefs at a single time is very unlikely, since once I locate a single belief relevant to the context, I stop searching for new ones. Augmented with the quantitative principle, this would be enough to explain why sometimes one fragment dominates another: the dominant fragment is more likely to be retrieved (quantity) and once retrieved, retrieval ceases (sleeping dogs).

But this does not seem like a particularly good explanation of what is happening in the studies by Shtulman & Legare, insofar as they observe competition between explanatory frameworks and thus presumably both are activated. We can also observe a tension between the diagnosis of representational fragmentation and the studies by Astuti, since she describes the two inconsistent explanatory frameworks of dualism as a foundational part of the development of the acquired-trait framework, even though the two disagree, whereas the metaphor of fragmentation would suggest the fragments are not related.

More generally, Bendaña & Mandelbaum's notion of belief is extremely permissive, in that they allow a broader swathe of representations to count as beliefs. This allows them to accommodate incoherence and even, as we've seen, extend it to almost all beliefs. But the downside of this approach is that we miss out on the different degrees and strengths of commitment we take to explanatory frameworks. The central question of explanatory coexistence is not just understanding when we believe and do not believe in the frameworks, but how they are sometimes endorsed, accepted, considered, utilized, and so on – how they are held, in other words, in an evolving series of attitudes. I've suggested that one way the fragmentation account could treat some of this variability is by appealing to probability of retrieval. But this is

insufficient for several reasons: likelihood of retrieval cannot be explanatory of differences once both frameworks are retrieved, it is in conflict with there being connections between inconsistent frameworks, and furthermore, it is itself unexplained.

The representationalist might reply: why think the attitudes we take are so complex as all that? In response, I point them to the debate woven throughout Astuti's paper: generations of anthropologists arguing about what it would take to really believe something, and fitting various attitudes and justifications to the circumstances. Or to the intuitive shades of difference between really endorsing something, using it in thought, considering it for a moment, and so on. So the very feature of this form of representationalism that allowed it to accommodate incoherence, its permissiveness about belief, is just what stops it from explaining explanatory coexistence: representationalism (of the Bendaña & Mandelbaum variety) is so permissive that almost everything counts as belief, and if that is so, then the idea of a framework passing in and out of being believed is hard to analyze.

On the second strategy, we would put forward a different version of representationalism that did not posit Persistence. This strategy is beyond the scope of this chapter: I'm not aiming to defeat all (or even any) types of representationalism. But since it follows on the previous thought, I note that if you want representationalism in the strong sense, that is belief-tokens really in the head, it is natural that you will end up with something close to persistence. This is because in order to capture enough representations that you can explain action, thought and behavior – that is, to be a pure representationalist and not make belief a rare phenomenon – you will have to loosen something about the notion of belief. Mandelbaum and coauthors choose to loosen the attitudinal component, but I suspect that the alternative is to loosen the representational component, and hence depart from representationalism of the strongest kind.

Now onto dispositionalism. Dispositionalists already struggle with incoherence of any kind, because a disposition to act in a way consistent with a belief that P is more or less a matter of actions cohering with beliefs. That is, we can tell what it means to act as if you have an orange in your bag, and to act as if you don't, by thinking of these actions as making sense given the belief. But nothing much makes sense given the belief that you both have and don't have an orange.

At this point, the dispositionalist has two strategies that can be used separately or together. These are relativizing beliefs to contexts, and minimalism about the significance of belief. Relativizing beliefs to contexts is just the move we've already seen in the case of fragmentation: roughly, if I believe there's an orange in my bag (relative to being asked explicitly), I can also believe there's not an orange in my bag (relative to trying to find it myself). This runs into a challenge with Partial Contextuality, the idea that explanatory coexistence involves at least partially overlapping contexts. This conflict can be mitigated by allowing for a more nuanced notion of context, perhaps, but it is impossible to narrow the context so much that two frameworks are brought up and conflict with one another in producing the very same behavior, as we see in the Shtulman & Legare paradigm.

For the second strategy, we might follow Eric Schwitzgebel (2002) in holding a minimal conception of belief, on his analogy with personality. On his view, belief is not a fundamental psychological kind but a helpful though superficial way of speaking about behavior and thoughts. If so, we might not be so troubled by difficulty in ascribing conflicting beliefs, either because this may just be an inherent limit on the usefulness of the category, or because the superficiality of the category itself allows us to ascribe “in-between” or sort-of believing in cases that don’t fit the mold perfectly. The more complex the bundle of dispositions are, the more we can imagine a case in which some of the dispositions pattern with one belief, and some with an inconsistent belief (as Schwitzgebel terms it, the more complex bundle is a dispositional stereotype). I have some sympathy with this strategy but it should not be our first resort. So for now, I set it aside.

I’ve discussed some difficulties for accommodating explanatory coexistence within existing frameworks of belief. These may not be fatal. But they lead us to ask: why is it so hard to place explanatory coexistence in these frameworks? I suspect it’s because these fairly extreme theories of belief either impose lots of consistency or very little - but to pose Gelman’s question we need to see how pressure for consistency is applied more in some cases than others.

### **3. Planning**

Let’s return to the beginning. Starting with a few cases of “double planning”, I’ll show that double planning is a kind of coexistence. A few key assumptions about what planning is and its relationship to having a plan allow us to formulate satisfying answers to the question: why do we double plan? I aim to extract these assumptions here, and apply their analogues to belief in the next section.

#### **3a. What is coexistence in planning?**

Double planning is when people prepare to act in ways that reveal incompatible assumptions. Like explanatory coexistence, this phenomenon involves a lack of coherence that may or may not be noticed or reconciled. Let’s look at a series of examples.

Azra is planning to go to nursing school. This means she is thinking through more specific sub-plans, such as: taking one of the required pre-requisites this semester, going to speak to an advisor about applications. Some of these sub-plans may be: fully enumerated (Going to talk to Dr. X on Wednesday) or incompletely enumerated (do something to fill out CV), and much of what it means to be “thinking through” sub-routines is just asking herself questions (should I try to get an internship?). In fact, she already counts as planning to go to nursing school once she starts asking these questions *as* questions about sub-plans. Thus she is planning even before she’s formed any sub-plans. She counts as planning just in virtue of asking the questions in the right way rather than holding in her mind a plan with the content: go to nursing school.

This is just a standard case of planning. To see where incoherence comes in, we’ll add in a contingency:



Bitra is planning to go to nursing school, if she scores over a 95% on her next exam, or else to join the Coast Guard. She plans just as Azra does for nursing school, and at the same time plans for the Coast Guard, thinking through sub-plans for both, sometimes at the same time.

Bitra has two sub-plans that cannot be enacted together, but is not incoherent since they are combined by a clear contingency. It's this contingency, that, when relaxed, give us double planning:

Celina is just like Bitra, except she does not have a clear idea of what would make her go to nursing school or the Coast Guard: perhaps something external, like test scores, or perhaps a decision based on her own values, once she has more time to think. For now, she is planning for both possibilities. If you asked her whether she could both go to nursing school and go to the Coast Guard, she would say no.

Celina is our first example of a real double planner. She is thinking and acting on two incompatible plans at the same time, in a way that is self aware. We can also imagine a variation in the case that is less self aware:

Destiny is just like Celina, except she is not sure whether it might be feasible to both go to nursing school and also join the Coast Guard.

This last two of this series of cases illustrates double-planning, which we are now in a position to define. Double planning is the activity of generating, evaluating, and otherwise considering multiple incompatible plans over the same period of time. These plans are incompatible in the sense that they cannot be jointly realized, but in some cases, the plans can be *rendered* compatible by means of a contingency.

Double-planning, like planning in general, is an activity or behavior, which in many cases may be entirely mental though in others involves external actions such as making a list, doing research, and so on. Further, planning is not reducible to having a plan. To see this, note that I can have a plan without being in the process of planning, such as when I have already finished planning yesterday. Conversely, I can be planning without having a plan, such as when I have just begun to consider a problem. But double-planning cannot be diagnosed when at least a partial plan is not present, since it is the plans which conflict.

### **3b. Why is there coexistence in planning?**

With this picture of double-planning in hand, let's turn to the question of *why* double-planning arises. I'll separate the answers into two families: explanations arising from limitations in computational capacity, and those that are independent of computational limits.

One source of explanations is that avoiding double planning would involve a burdensome amount of monitoring. A simple way of spelling out this point is in terms of computational limits: we just could not always resolve our plans for coherence, and engaging in monitoring even substantially

under the limit of what technically could achieve would take away resources that are needed elsewhere. But excessive reconsideration might be bad even irrespective of limitations. On this point, Sergio Tenenbaum (2020) writes:

Quite often reconsidering is not too costly, and for (almost) any particular plan, abandoning just this plan will not undermine more general ends. But reconsidering in too many cases can have a devastating cumulative effect: were I to reconsider my intentions at every permitted opportunity, I would forgo pursuing many of the ends I care about. And were reconsideration to lead me to revise plans often enough, my life would be a pathetic alternation of momentary or soon-to-be-abandoned pursuits. On the one hand, no particular intention must “resist reconsideration”; the requirement not to reconsider too much applies only to the total set of one’s intentions. On the other hand, we can only satisfy this requirement if we avoid reconsidering particular intentions, none of which we are required not to reconsider.

The idea that too-frequent reconsideration would lead to a life of abandoning plans rather than sticking with them is a problem based not in limitations but the supposition that often enough, the only reason I have to continue on with my pursuits is that I am doing so – as soon as I ask myself why I should do this rather than that, I may be inclined to rationally alter my plans. At least in this paraphrase, this problem with reconsideration would seem to be dissolved by higher-order reasons of the right kind to stick to our projects, whatever they are. But in the absence of that or other reasons that tie us to the mast, Tenenbaum’s reasoning shows that avoiding excessive reconsideration is a necessary diachronic strategy for averting a descent into a mire of transient goals. Connecting this to double-planning just requires the premise that the consequence of reconsideration is typically increased coherence, and conversely, lack of reconsideration allows contradictory plans to flourish. This premise might be motivated, for instance, by the idea that reflection in general (of which reconsideration is a subtype) aims at coherence.

Bratman et al (1988) raise a related concern about reconsideration, which on their architecture enters in at the level of an override of the “filter” which limits the plans under consideration based on compatibility (among other things). The filter override mechanism can, for instance, allow a plan to be reconsidered despite conflicting with a current plan:

An agent’s filter override mechanism must be carefully designed to embody the right degree of sensitivity to the problems and opportunities that arise in her environment. If the agent is overly sensitive, willing to reconsider her plans in response to every unanticipated event, then her plans will not serve sufficiently to limit the number of options about which she must deliberate. On the other hand, if the agent is not sensitive enough, she will fail to react to significant deviations from her expectations.

On Bratman et al’s picture, the option set that we choose from has already passed through a (leaky) process that screens for compatibility. In the case of double-planning, some forms of double-planning may be the consequence of leaks in the filter itself. A second route for

incoherence to make its way in is through the override, though this would immediately lead the two inconsistent options to be considered together. But the result of this consideration need not be coherence. The planner might instead persist in incoherence, mulling it over, or leave the issue to be resolved later. As in Bratman (1992), the picture of coherence at work here is one in which we tend towards coherence over time when things go well, rather than maintain it fully at any one time.

Avoiding excessive reconsideration gives us a reason to expect double-planning as a side-effect. But we can also locate a more direct rationale. The examples of Celina and Destiny differ from Bitá in that Bitá has identified and committed to a precise contingency, whereas Celina and Destiny will resolve the contingency later or not at all. There may be a benefit to resolving contingencies when sufficient evidence has been acquired for the correct choice to be made, or minimally, for a measure of stability to obtain. For instance, she might be about to work a summer job in a hospital or have an appointment with a guidance counselor about her career plans. In that case, it might do to postpone settling on a contingency, either in order to save mental effort, or against the background of the argument given by Tenenbaum, even an agent of unlimited capacity who did not always reconsider would have to take care to not settle on the wrong contingency and be stuck with it until her next reconsideration.

More generally, consider Bratman's (1992) set of pressures on acceptance of options in planning, among which he counts: simplicity, error asymmetries, social cooperation, moral or relational pressures, and pre-conditions for practical reasoning. This list is meant to capture sources of constraint on acceptance of options that do not apply to belief, a point which will be of interest later. But for now, each of these reasons in the right context may favor double-planning. Double-planning could simplify a problem when contingencies are complex, could help avoid costly error when the risk of having no plans vs conflicting plans is relatively high. Social pressures provide a different kind of rationale for double-planning: I might, by double-planning, be able to plan together with two different associates at the same time rather than choosing between them. Likewise, relational and moral reasons might drive me in both directions simultaneously, and lastly, a precondition for practical reasoning at all might be that I am so critical of my own plans that I can't get started. For this reason, I might need to double-plan, at least if I want to be sufficiently ambitious.

Taking stock, double-planning is a structurally similar phenomenon to explanatory coexistence. Both involve working with frameworks that are inconsistent. But where explanatory coexistence fits somewhat uneasily with theories of belief, double-planning is a core part of accounts of planning. We've seen that a fairly wide range of rationales might be given for double-planning, some applying to agents with certain limitations and others to any agent whatsoever. In the next section, I'll put forward an account of belief that makes a close analogy with planning, which will allow us to give a similar analysis of what explanatory coexistence is and why it obtains.

#### **4. Coexistence on the planning theory of belief**

We are now in a position to put these pieces together. As we saw, planning is an activity that involves computation, often alongside other actions both mental and physical, as we figure out

what we should do. It is not reducible to having a plan. The planning theory of belief borrows this general structure. We start, not with belief, but with believing, which I understand as an activity that involves computation, often alongside other actions both mental and physical, as we figure out what is the case. The planning theory takes believing, not belief, to be primary.

When I am planning to go to the store, I often do this by considering various subplans of ways to go to the store. It is in most cases enough to be planning to go to the store that I be considering such subplans – nothing further is required, though defeating conditions (such as being engaged in pretense) might get in the way. Likewise, when I am believing that witchcraft causes disease, I often do this by considering various sub-questions and theories under that umbrella, such as evaluating the claim that this person's disease was caused by this act of witchcraft, or asking myself whether measles is caused by this or that kind of witchcraft.

Thus we have an activity, believing, which is directed towards a theory or more minimally, a single proposition. Belief, as a state, is understood in a derivative sense: my set of beliefs at any given point in the process of thought is just that set of propositions that best captures *where I am in the process of thinking*. In many cases, this will be stored in a declarative representation, just as in many cases of planning, I have a current plan representation that more or less accurately captures my current state of mind about how to achieve my end. But consider a case in which my thoughts have shifted substantially in the sense that I am asking myself questions that conflict with my previous understanding, which is more explicitly represented. In this case, we would say that my beliefs are not identified with what I represent explicitly, because overall, my current state of mind about the issue suggests a different position. So beliefs as a state at a time are an abstraction that corresponds to my current state of mind (in fact, there are many possible abstraction functions depending on context and demand, see Aronowitz (forthcoming) for an extended discussion). This is quite close to the dispositionalist understanding of belief. Believing, on the other hand, is a real computation over representations – just not exclusively or even necessarily including beliefs. This shares something with the representationalist in making believing a fundamental, operational psychological category. In that sense, the planning theory is a middle ground between the two extremes.

Let's at last turn back to explanatory coexistence. On the planning theory, the competing frameworks are representations used in believing but need not themselves be beliefs. This is on analogy with the way many types of representations other than current, active plans enter into planning – for instance, planning architectures often have a plan library, an archive of past problems and solutions, that are searched as part of the process of planning. Just like a plan library contains past or even never-endorsed plans that are not what we are currently planning, long term memory contains past or even never-endorsed representations that are not we currently believe but still can be used in believing. This allows the planning theory to leave a crucial psychological question open: in each case of explanatory coexistence or in general, to what extent are the competing frameworks adopted? Are they fully believed, merely used, or something in between?

So much for accommodating the idea of coexistence. More interestingly, some but not all of the rationales surveyed for double-planning will correspond to rationales for explanatory coexistence. The analog of not reconsidering plans too often, in the case of believing, would be a principle of not reopening questions too often. Just like plans need to be settled to some degree for us to live a life with large-scale practical projects, some questions might need to be settled sufficiently to allow us to pursue large-scale epistemic projects. Then, we might be better off with explanatory coexistence because to resolve competing frameworks for coherence would require too much reconsideration. Somewhat surprisingly, this does not seem like the same kind of danger, perhaps because the typical explanatory coexistence examples are frameworks that persist for years and decades. That is, we would surely have an opportunity to reconsider them, so this rationale is not convincing.

However, there is another cost of resolving beliefs through reconsideration. When operating with the explanatory frameworks of witchcraft and germs as causes of illness, if we were to resolve for coherence, we would either need to come up with a complex unifying theory that allowed both to operate alongside one another, or to choose which one to retain and which to jettison. The former option, creating a unifying theory, may be either hard to square with the evidence, or just elusive. That would leave us with the need to pick just one.

Again drawing on the parallel with planning, we can sketch two distinct reasons for coexistence to continue. First, if I expect to gain more evidence later about which framework is more apt, I have reason to delay resolution for now. Second, reasons of social or moral connections might confer distinct advantages on each of the theories: for instance, social cooperation might require fluency in a commonly-endorsed framework, where each framework is more commonly endorsed in a different subset of the community I interact with. Returning to Astuti's participants, her understanding is more or less that Vezo adults both endorse dualism and are pulled to speak and act in ways that accord with the conflicting, fully-acquired-traits framework, by reasons of social cooperation. She connects this to the prominence of non-kinship-based social structures in the Vezo community: to respect these structures is in part to avoid drawing too much attention to biological descent. But there are presumably practical and epistemic reasons to favor the dualistic conception as well, for instance that it fits with observations about biological inheritance. We can see that the argument already given by Astuti fits a pattern familiar from the case of planning.

There is much more work to be done, both empirical and philosophical, in understanding why explanatory coexistence occurs. But I have said enough to establish that the planning theory of belief sets up a straightforward and fruitful way to understand these advantages. Unlike representationalism, the planning theory allows us to separate representations used in believing from beliefs proper and thus delineate different roles for these representations in thinking. Unlike dispositionalism, the planning theory concerns itself directly with computation over representations, and so can dig into questions of tractability and efficiency, which as we've seen are highly relevant for explaining coexistence.

## **5. Conclusion**

Explanatory coexistence is one of many kinds of incoherence in believing. While incoherence of this kind might be considered an error, or even conceptually impossible, by philosophers, psychologists and social scientists have made the case that explanatory coexistence is widespread and if not clearly rational, then at least not particularly aversive. This raises a question: why does coexistence develop, and when and why is it maintained or resolved? In this chapter, I've suggested that this question is hard to address on two pictures of belief, representationalism and dispositionalism. Then, I made the case that a similar sort of coexistence in planning, double-planning, is less puzzling: it is easier to describe and link to theories of planning, and we have some insight into why it should occur. The reason this difficulty is dissolved in the case of planning, I argued, is not because of a deep asymmetry between believing and planning, but because we have gone wrong in our basic understanding of what belief is. When we restore symmetry between belief and planning by centering our account on the activity of *believing*, rather than belief states, coexistence can be seen more clearly and we can take the first steps towards answering the question of why it occurs when it does. Elsewhere, I defend the planning theory of belief on other grounds, but in the present paper, we can only draw the modest conclusion that it has one advantage over its competitors.<sup>3</sup>

## References

- Aronowitz, S (in press). A Planning Theory of Belief. *Philosophical Perspectives*.
- Bendaña, J., & Mandelbaum, E. (2021). The fragmentation of belief. In Borgoni, C., Kindermann, D., & Onofri, A. (Eds.). *The fragmented mind*. Oxford University Press.
- Bratman, M. E. (1992). Practical reasoning and acceptance in a context. *Mind*, 101(401), 1-15.
- Bratman, M. E., Israel, D. J., & Pollack, M. E. (1988). Plans and resource-bounded practical reasoning. *Computational intelligence*, 4(3), 349-355.
- Elga, A., & Rayo, A. (2022). Fragmentation and logical omniscience. *Noûs*, 56(3), 716-741.
- Legare, C. H., & Gelman, S. A. (2008). Bewitchment, biology, or both: the co-existence of natural and supernatural explanatory frameworks across development. *Cognitive science*, 32(4), 607–642. <https://doi.org/10.1080/03640210802066766>
- Quilty-Dunn, J., & Mandelbaum, E. (2018). Against dispositionalism: Belief in cognitive science. *Philosophical Studies*, 175(9), 2353-2372.
- Ryle, G. (1949). *The Concept of Mind*.
- Schwitzgebel, E. (2002). A phenomenal, dispositional account of belief. *Noûs*, 36(2), 249-275.
- Schwitzgebel, E. (2001). In-between believing. *The Philosophical Quarterly*, 51(202), 76-82.

<sup>3</sup>Many of these ideas were developed in conversation with Tania Lombrozo, for which I am deeply grateful. Thanks also to Reza Hadisi, Harvey Lederman, Nic Porot, Julia Staffel, and audiences at CU Boulder and LSE for feedback on a related talk.

Shtulman, A., & Legare, C. H. (2020). Competing explanations of competing explanations: Accounting for conflict between scientific and folk explanations. *Topics in cognitive science*, 12(4), 1337-1362.

Sommer, J., Musolino, J., & Hemmer, P. (2023). A hobgoblin of large minds: Troubles with consistency in belief. *WIREs Cognitive Science*, 14(4), e1639. <https://doi.org/10.1002/wcs.1639>

Tenenbaum, S. (2020). *Rational Powers in Action: Instrumental rationality and extended agency*. Oxford University Press.

Van Fraassen, B. C. (1980). *The scientific image*. Oxford University Press.