Abstract

This chapter motivates a fragmentationist research program by identifying a cluster of problems that such a research program is better positioned to address or resolve than a unified model—all instances of the phenomenon of subjects having information that's available to them for some behavior-guiding purposes but that isn't available for every purpose. It also identifies some of the challenges and research questions that the fragmentationist program will need to address and where the space of possible answers is not yet well charted. One cluster of such problems is about how to construct fragmented models of belief, and another cluster is about normative questions that arise concerning the rational evaluation of fragmented beliefs and believers.

Keywords

fragmentation, belief, credence, rationality, updating, decision, inconsistent belief

4

Fragmented Models of Belief

Andy Egan

1. Introduction

This chapter is primarily an advertisement for a research program, and for some particular, so far under-explored research questions within that research program. It's an advertisement for the program of constructing fragmented models of subjects' propositional attitudes and theorizing about and by means of such models. I'll aim to do two things: first, to motivate a fragmentationist research program by identifying a cluster of problems that such a research program is well positioned to address or resolve, and second, to identify what I take to be some of the challenges and research questions that the fragmentationist program will need to address and where the space of possible answers is not yet well charted.

Many of the tools that philosophers standardly use to model and theorize about belief presuppose a *unified* picture of subjects' doxastic states. Consider, for example, the theorist who represents a subject's doxastic state with a set of possible worlds (Lewis 1979; Stalnaker 1984; Braddon-Mitchell and Jackson 2007). The role of this set of worlds is to model how things are, according to the subject, by identifying the possibilities that the subject treats as live and those they rule out. This set of worlds is then the object that the theorist reaches for to explain all of the subject's behavior (or rather, all of the subject's behavior that's susceptible to belief/desire explanation. This qualification will often be omitted hereafter). Since standard possibility-

carving models associate each subject with a *single* set of worlds, the model is unified in the sense that there's a single object that we appeal to in order to explain all of the subject's belief-governed behavior. (This is a picture that fits naturally, as Yalcin 2018 points out, with Ramsey's 1931 characterization of belief as a 'map by which we steer.' The set of worlds we associate with a subject to model their doxastic state serves to characterize the map by which the subject steers—the map that guides the subject's belief-governed behavior.)

Other standard tools for modeling doxastic states, such as associating a subject with a credence function, or with a (consistent and logically closed) set of propositions, share this feature. On standard implementations of these sorts of models, a single subject (at a time) is associated with a *single* credence function, or set of propositions. And that single object is the thing that the theorist will reach for in order to account for all of the subject's belief-governed behavior.

A subject's beliefs, on this sort of model, are all 'always on.' There's a single object that's being used to model the subject's beliefs, and that object is the one that's appealed to in accounting for all of the subject's belief-governed behavior. And so if the object that we use to account for *this* bit of the subject's behavior (or *this* disposition to behavior) encodes a belief that *P*, then so must the object that we use to account for *that* bit of, or disposition to, behavior.[1] So

[1] This is straightforward in synchronic cases, where the two bits of behavior occur at the same time (when I walk and chew gum, or drive and text (don't do this!), or perform delicate surgery and talk about politics (probably best not to do this either)). It's more complicated in the diachronic case, where the two bits of behavior happen at different times. See later in this chapter for a bit more discussion of this issue.

these are all *unified* models, encoding the assumption (or the idealization) that all of the information that's brought to bear in the production of any aspect of the subject's belief-governed behavior is brought to bear in the production of all of it.

The fragmentationist program is to move from unified models to models that characterize subjects' doxastic states as fragmented. Fragmented models of belief are ones that do not encode the assumption that all of a subject's beliefs are 'always on' and that all of the same information is brought to bear on the production of all of a subject's belief-governed behavior. To put it another way: these are models where subjects don't always steer by the same map, don't guide all of their behavior with the same representations, and don't bring all of the information they possess to bear on all of their belief-governed behavior.

The project of spelling out that kind of fragmented model turns out to be rather complicated.

One reason for this is that it's not totally straightforward how to build a framework for theorizing about fragmented belief—there are questions about what to add to or change about our familiar unified models that don't have obvious answers. I'll make a few suggestions about possible answers to some of these, but I'll be more interested in drawing attention to some of the choice points.

Another reason is that when we move to a fragmented model of belief, a number of new normative questions arise (perhaps along with some new metaphysical questions), and the answers to those aren't obvious either. I'll draw attention to some of these.

In the remainder of this introductory section, I'll preview from high altitude the kind of motivation for moving from unified to fragmented models that I'll spend Section 2 elaborating on. I'll then provide a roadmap of the rest of the chapter. And then we'll get on to business.

One difficulty with standard unified models is the kind of logical closure and/or probabilistic coherence that's built into them. If we're representing beliefs with a single set of worlds and we've got a subject who believes that *P* and believes that if *P* then *Q*, then we can't help but have a subject who believes that *Q*. Responding to this sort of difficulty has been a central motivation for fragmented pictures. But the motivations for fragmentation are undersold if we think of fragmentation as a merely defensive move in the project of logical omniscience apologetics, and in order to avoid giving that impression I will not focus on this motivation for fragmentation in what follows (but see for example Braddon-Mitchell and Jackson 2007; Elga and Rayo (Chapter 1 in this volume); Fagin et al. 1995; Field 1986a, 1986b; Soames 1985, 1987; Speaks 2006; Stalnaker 1984, 1991; and Yalcin 2018).

Another difficulty—the one I will be spending a lot of time on in what follows—arises from unification as such (rather than from the fact that the particular unified belief representor we're using is one that enforces closure and the like). If we've got just a single object (credence function, set of worlds, etc.) representing a subject's doxastic state, then we must appeal to that object in all of our belief/desire explanations of the subject's behavior. If an action or disposition to action can't be explained by appeal to that object, then it can only be explained in non-doxastic terms (see Section 2.3 for support for the claim that this really is a problem for unification as such and that dropping closure and consistency requirements from unified models doesn't help to address it).

(Extended parenthetical aside: here I am suppressing complications about the diachronic case. Of course subjects' beliefs change over time. And so of course a theorist who's working with a unified model won't in general need to appeal to the same credence function, set of worlds, etc. when they're explaining two bits of behavior that occur at different times. Instead,

they'll want to appeal to a particular credence function (set of worlds, etc.) C1 to explain the bit of (or disposition to) behavior at (earlier) *t*1, and then to a credence function, etc. C2 that's the result of applying some plausible series of updating procedures to C1 in order to explain the bit of (or disposition to) behavior at (later) *t*2. So in general there will be two strategies for making trouble for unified models: first (and this is what will mostly happen in what follows), identify examples of contemporaneous behaviors or dispositions to behavior that aren't happily accounted for by appeal to the same credence function, etc. Second, identify examples of behaviors or dispositions to behavior at different times, such that the sort of credence function (etc.) that would be well positioned to account for the later behavior (disposition) isn't one we can get to by any plausible series of updates from the sort of credence function (etc.) that would be well positioned to account for the earlier behavior (disposition). I think there are plenty of the second kinds of cases too, but since they're a bit more dialectically complicated, I will mostly stick to the first kind in this chapter. End of parenthetical aside.)

If we find a pattern of (dispositions to) behavior such that part of it would be well explained by a credence function C1 and another part of it would be well explained by a different, incompatible C2, we will need to choose. One part of the overall pattern of behavior will get belief/desire explanation, and the other part will need to be written off as not actually belief-governed and explained in some other terms.

Generalizing a bit, the central problem with unified models is that there are lots of subjects, both actual and merely possible, who display patterns of behavior such that: behavior type K1 would be happily explained in terms of a credence function, set of worlds, etc. with property P1; behavior type K2 would be happily explained in terms of a credence function, etc. with property

P2; no unified doxastic representor has both P1 and P2; and it's highly theoretically desirable to be able to offer belief-based explanations of both K1 and K2.

The role of the next section will be to fill in some examples of this kind of phenomenon. Section 3 introduces the bare bones of a fragmentationist response. Section 4 takes up complications that arise when filling in the details of a fragmented model of belief. In Section 5, I survey some normative questions (and briefly allude to some metaphysical ones) that arise from the move to a fragmented framework.

2. Troublemaking Phenomena

In this section I'll canvass three types of phenomena that do not sit well with, and are not happily modeled with or explained by appeal to, unified accounts of belief. They are the distinction between recognition and recall, the distinction between ignorance and failure to bring to bear, and the phenomenon of inconsistent belief.[2]

2.1. Recognition vs. Recall

[2] For discussion of similar phenomena, and an at least superficially different proposed response, see Eric Schwitzgebel's work on 'in-between belief' (Schwitzgebel 2001, 2002). I think it's an interesting question just how different Schwitzgebel's account is, ultimately, from a fragmented framework (very many of the same questions will arise in filling out the details of a characterization of a subject's total doxastic state in Schwitzgebel's framework), but I won't take that up here.

Consider Lucille's beliefs about the film *The Sound of Music*, as manifested by her dispositions to respond to various questions about it.

Lucille is disposed, when asked

(1) What was the name of the youngest von Trapp child?

to hem and haw for a moment and then to respond, 'I don't know.'

She is also disposed, when asked

(2) Was the youngest von Trapp child's name 'Gretl'?

to nod emphatically and say, 'Yes, of course.'

Here is a puzzling feature of this sort of case: the information required to answer the two questions is exactly the same. In this case, it's the information that the youngest von Trapp child's name was 'Gretl.'

(I trust that this general pattern, in which some piece of information is easily *recognized* as true when one is presented with it explicitly but not easily *recalled* in response to an open-ended question, is familiar to the reader. Just which questions people have these sorts of patterns of dispositions with respect to is highly interpersonally variable. Maybe you have (or had, until just now) the same dispositions as Lucille with regard to these particular questions. Maybe not. But in any event, it should be easy to come up with a case that works for you.) (See also, for example, Stalnaker 1991; Egan 2008; Bendana and Mandelbaum, Chapter 3 in this volume; Kindermann and Onofri, the Introduction to this volume; and Elga and Rayo, Chapter 1 in this volume. See Kindermann, Chapter 8 in this volume, for analogous cases having to do with conversational presuppositions.)

Here is the awkward question for theorists working in unified frameworks: Before anybody asked Lucille any questions, what did she believe? What, for example, was her credence that the

youngest von Trapp child's name was 'Gretl'? What distribution of credence should we assign to her in order to explain her overall pattern of dispositions to behavior?

There is no happy unified answer to these questions. In order to explain her hemming, hawing, and 'I don't know'-ing in response to (1), we should assign to her a credence function that's widely spread out over the possible answers—one that does not assign a very high credence to the proposition that the youngest von Trapp child's name was 'Gretl' or to any other specific answer. In order to explain her confident 'Yes, of course' response to (2), we should assign to her a credence function that's highly concentrated on (assigns high credence to) a particular answer—the correct one, that the youngest von Trapp child's name was 'Gretl.' No single credence function is well positioned to represent Lucille's overall doxastic state, which gives rise to both of these dispositions.

(It should be clear that the situation is no better for other sorts of unified models—no single set of worlds, or single consistent and closed set of propositions, will do the trick either.)

Here is a natural thought to have about the case, and something that it would be nice to be able to say about it: what's happening with Lucille is that the information about Gretl's name is there, stored in her mind somehow, but it's not always available for behavior guidance. It's easier to 'call up' in some circumstances than others. That's why she is disposed to hem and haw in response to (1) and answer with a confident 'yes' to (2)—because the situation in which she is prompted with (2) is one that makes the information available, but the situation in which she is prompted with (1) is not.

But this is not something that a unified model allows us to say. In a unified probabilistic model, Lucille has, at any given time, a single credence assignment to the proposition that the youngest von Trapp child's name was 'Gretl.' A high credence puts us in a good position to

predict and explain the 'yes' in response to (2), but not the hemming and hawing in response to (1). A low credence puts us in a good position to predict and explain the hemming and hawing in response to (1), but not the 'yes' in response to (2). In a unified model using sets of worlds, either Lucille's belief worlds will be restricted to worlds in which the youngest von Trapp child's name was 'Gretl' or else they won't. The set that includes only 'Gretl' worlds will be well positioned to account for Lucille's confident 'yes' in response to (2), but not the hemming and hawing in response to (1). The set that includes some non-'Gretl' worlds will be well positioned to explain the hemming and hawing in response to (1), but not the 'yes' in response to (2). In a unified model using sets of propositions, either the proposition *that the youngest von Trapp child's name was 'Gretl'* will be in the set that represents Lucille's beliefs or it will not. And again, each option leaves us well positioned to explain one disposition to respond but ill positioned to explain the other.

There is no room in such models for variable accessibility. The information (or the high credence) is either encoded in the object we use to represent the subject's belief state or it isn't. There's nothing in these frameworks that can represent the ways in which not all of a subject's beliefs are 'always on'—how some of the information that a subject can use to guide their behavior in some circumstances isn't available for behavior guidance in *every* circumstance.

2.2. Ignorance vs. Failure to Bring to Bear

Let's move on to another troublemaking distinction, which unified pictures are also ill suited to model: the difference between *ignorance* and *failure to bring to bear*. I'll again illustrate with an example.

(This case was offered to me as an account of actual events by a very cognitively sophisticated and capable friend from graduate school during a discussion of fragmentation. I will put the case in the first person to protect their identity.)

I'm watching a movie on my TV during a thunderstorm. The power to the house goes out, and the TV goes dark. I think to myself, 'Oh, I can't watch the movie anymore—I'll just go check my email,' and head upstairs to go turn on my computer.

My plan (as you will probably have predicted) doesn't work. When I flip the switch on the computer, it doesn't turn on—if the power to the house is out, the computer isn't going to work any better than the TV. And my reaction after I flip the switch on the computer and nothing happens will suggest that there is a clear sense in which I already knew this—I may slap my forehead, look around in embarrassment to see if anyone has noticed my mistake, tell the story later to my friends who are writing papers about fragmented belief, make them promise not to reveal my identity when they use the example in a paper, etc.

This sort of case, in which there's some piece of information such that (a) we want to say that the subject knows it—we don't want to attribute ignorance—but also (b) the subject doesn't bring it to bear on a particular piece of action or deliberation, is ubiquitous.

My going upstairs to check my email looks like a piece of deliberate, intentional, goal-directed behavior, which we'll want to explain in terms of my beliefs and desires. It's not a reflex, or a hardwired response, or an automatic subroutine activated by some stimulus, or anything like that. But to explain my going up to check my email on my computer, we'll want to reach for a belief state according to which the computer is likely to work. And so we won't want to reach for one according to which (i) the electricity to the house is out, (ii) if the electricity to the house is out, then the electricity to the computer is out, and (iii) the computer needs

electricity to work. But it also seems absurd to deny that I believe all of those things. To explain a bunch of other things I'm disposed to do, in other circumstances, we'll want to reach for a belief state that *does* include all of that information (perhaps my embarrassed forehead-slapping, and perhaps my disposition to correctly answer, in many circumstances, questions like 'If a power outage cuts power to the TV, will the computer work?' and 'Why not?').

Again, there doesn't seem to be a single unified belief state of any of the familiar types that's well positioned to explain all the stuff about me that we want to explain. There's no set of binary on/off beliefs, and no single credence function, that's well positioned to explain both my going upstairs to check my email and my disposition, for example, to say, 'Yeah, duh' when asked, 'Does your computer need electricity to work?'

Here is a natural thought about how to describe my doxastic situation that might point to a way to change our models to accommodate the phenomenon: we don't always bring all of our beliefs to bear on any particular bit of deliberation or action. I've got the right beliefs about how computers work and how power outages standardly affect whole houses all at once. It's just that, when I'm deciding what to do after the TV stops working, I'm not bringing those beliefs to bear on my decision-making.

2.3. Inconsistent Belief

In a famous example from 'Logic for Equivocators' (1982), David Lewis describes his earlier pattern of belief about the relative positions of the train tracks and Nassau Street in Princeton. Lewis says that at one time he believed: (a) that the train tracks ran roughly north–south, (b) that Nassau Street ran roughly east–west, and (c) that the tracks and Nassau Street ran roughly parallel. If we resolve the context-sensitivity of 'roughly' so that the standards aren't too loosey-

goosey,[3] these three are inconsistent. (In the remainder of this chapter, I'll drop the 'roughly' from Lewis's example as it's not essential to the point and makes presentation more complicated.)

Let's fill in the case a bit with a behavioral/dispositional backstory that would make this pattern of belief attribution attractive.

Suppose that this is how Lewis is disposed to behave:

- When walking along Nassau Street, he is disposed, when asked where the North Pole is, or which way is north, to point in a direction that is perpendicular to the street (and toward where he thinks, for example, Ottawa, or Montgomery NJ, is).
- When on the train, he is disposed, when asked where the North Pole is, or which way is north, to point along the route of the tracks (in the direction that he takes Princeton station to lie).
- All the time, Lewis is disposed, when he's either on Nassau Street or by the train tracks and asked about the location and direction of the other, to say that they run approximately parallel over that way (pointing perpendicular). And he's consistently inclined to follow a 'walk at right angles' strategy to get from one to the other.

Philosophers working with a unified model of belief will be hard pressed to accommodate this case in a satisfactory way.

What should we say that Lewis believes about the geography of Princeton? His behavioral dispositions support attributing to him each of the beliefs Lewis reports himself as having had: that the tracks run north–south, that Nassau Street runs east–west, and that the tracks and Nassau

[3] As Lewis instructs us to do—he stipulates that he means 'to within 20 degrees.'

Street run parallel. His disposition to point parallel to the direction of the tracks when asked which way is north would be well explained by a belief that the tracks run north–south. His disposition to point perpendicular to the direction of Nassau Street would be well explained by a belief that Nassau Street runs east–west. And his disposition to follow a 'walk at right angles' strategy to get from one to the other would be well explained by a belief that the tracks and Nassau Street run parallel.

But we can't include all three in a unified representation of Lewis's total doxastic state. There are no worlds in which all three are true, so there's no non-empty set of worlds with which to model such a belief state. There is also no consistent set of propositions that includes all three, and no coherent credence function gives high credence to all three (at least none that includes, or gives high credence to, all three propositions and also some geometric stuff that Lewis probably also believed).

One thing we could do, in response to this kind of case, is to lift some of our consistency, coherence, or possibility constraints on our unified models of belief. We then allow for *in*coherent credence functions, or sets of propositions that aren't consistent or logically closed, as representors of subjects' doxastic states. (The analogous move is harder for the possible worlds theorist, but one option would be to continue to represent a subject's doxastic state with a set of worlds while allowing that the set may include some *im*possible worlds, in which some contradictions are true.) This then allows us to add all three jointly inconsistent beliefs to our single unified representation of Lewis's doxastic state.

One reason that we might not like this sort of move is that we might have some fancy theoretical motivation for not countenancing these sorts of unified but inconsistent models for doxastic states. (Lewis himself, as well as fellow fragmentationist founding father Robert

Stalnaker, both have this sort of motivation, since they don't want to countenance impossible worlds; see for example Lewis 1986 and Stalnaker 1984.) There is much to be said about this, but we don't need to say it in order to see that this response is not terribly attractive. It's not terribly attractive, most importantly, because *just* changing our model so that it allows for unified representations of inconsistent belief doesn't actually help very much.

Here is why: what inclines us to attribute the inconsistent pattern of belief to Lewis is not that he acts, all the time and with respect to all of his behavior, like somebody who thinks that Princeton is put together in a geometrically impossible way. For one thing, it's not at all clear just how one would have to act, in order to act like somebody who believes that Princeton is put together in a geometrically impossible way. (How, for example, would one have to act in order to act in a way that would satisfy one's desires if *P* and not-*P* were true?) But even to the extent that we *can* make sense of this, it's not how Lewis acts. (And more importantly, it's not how the very many actual, non-fictionalized people act in the very many non-fictional cases that, like Lewis's, motivate the attribution of inconsistent belief.) He doesn't, for example, always and everywhere throw up his hands, or choose directions at random, or read up on fancy non-standard geometries or dialetheist logics when trying to navigate from place to place in Princeton.

What we see in Lewis's example as elaborated, and in the many real-world cases that Lewis's example brings to mind, is not a uniform pattern of peculiar, inconsistency-driven behavior, but rather two distinct patterns of behavior, each of which suggests a particular consistent picture of the geography of Princeton, where the two pictures suggested by the distinct patterns of behavior aren't consistent with each other. Lewis in the example (like actual people in analogous situations) does not uniformly, with respect to all of his behavior, act like a believer of

P and not-P (even to the extent that we can get a grip on what this would be), but instead acts in some circumstances like a *P* believer and in others like a not-*P* believer (for example, in one sort of circumstance, he's disposed to perform actions that would satisfy his desires if *P* were true, and in a different sort of circumstance he's disposed to perform actions that would satisfy his desires if not-*P* were true). That's the sort of pattern of behavior that would make us inclined to attribute inconsistent belief, and it's (a) very common in actual believers, (b) not happily accommodated by unified pictures of belief on which subjects' beliefs are always consistent, and, importantly, (c) also not happily accommodated by unified pictures of belief that allow for inconsistency.

This fact—that moving to unified models that don't enforce consistency or closure (or analogous constraints on probability distributions) doesn't actually give us models that are properly responsive to the troublemaking phenomena—is important. It's important because this kind of move is at least superficially attractive in response to lots of different kinds of objections to the sorts of unified, closure-and-consistency-enforcing models in wide use in philosophical accounts of belief. Faced with problems for such models, it's natural to wonder whether it's the unification or the enforcement of closure and consistency that's making the trouble (or whether it's the conjunction, and we could alleviate the pressure by dropping either one). The above discussion aims to give some general reason to think that it really is unification that's making the trouble, and dropping the closure and consistency requirements won't give us models that are better suited to addressing the troublemaking phenomena.

The ways in which moving to unified but inconsistent models fails as a response to these sorts of cases point us toward a more promising way to revise our models to accommodate the troublemaking phenomena (not terribly surprisingly, toward the way that Lewis thought we

should revise our models). This is to give up on unified representations of subjects' doxastic states and to move to a fragmented framework instead. What this looks like, and further questions that arise once we've decided to move to a fragmented framework, will be the subject of the remainder of this chapter.

Summing up before moving on: there are lots of phenomena that make unified models of subjects' doxastic states problematic. These phenomena are not recherché, not science-fictional, and not fancy theoretical problems that are artifacts of working with possibility-carving models. There are patterns of behavior that we observe in actual believers, which we'd like to say are belief-governed, and which we'd therefore like to be in a position to offer belief-based explanations for, but which are not happily explained in unified frameworks that treat the task of characterizing a subject's doxastic state as the task of characterizing a single, unified map by which the subject steers. We would be well served, if we want a framework for theorizing about belief that can accommodate these phenomena, to adopt a framework that allows us to represent *fragmented* systems of belief.

It's probably worth emphasizing at this point that the claim isn't that there are no phenomena such that unified models are theoretically useful for theorizing about them, or that there are no domains in which inquiry is best pursued using unified models. Instead, the point that I'm concerned to press here is that there are important and interesting doxastic phenomena that are not profitably explored using unified models, and that this shows that we shouldn't think that unified models get at the whole truth about, or the deep fundamental nature of, belief. There is a market for fragmented models as a way of productively theorizing about some phenomena that unified models are ill suited to account for—perhaps because fragmented models provide a better interpretationist framework to work in (see Williams forthcoming), perhaps because they

are better at characterizing the actual (and possible) structures of subjects' belief-underwriting representational apparatus (see Bendana and Mandelbaum, Chapter 3 in this volume), depending on what kind of theorizing about belief one goes in for.

In the remainder of this chapter, I'll take up the complications and research questions that arise once we decide to take this route.

3. Response: Fragmented Belief

Let's stick with Lewis for a bit, since the case is convenient for setting up the fragmented picture of belief as a response to the troublemaking phenomena.

The fragmentationist proposes that we don't attribute to Lewis a single set of inconsistent beliefs—we don't model his doxastic state with a single incoherent credence function, for example, or with a set of impossible worlds, or with a single inconsistent set of sentences, or propositions, or mental representational items. Instead, we attribute to him two distinct sets of consistent beliefs. And we modify the Ramseyan metaphor in a way suggested by Seth Yalcin (2018): we think of belief not as a (single) map by which we steer, but as an atlas from which we select the maps by which we steer. We don't treat Lewis as somebody with a single inconsistent map. We treat him as somebody with two maps, one of which he steers by when he's on Nassau Street, the other of which he steers by when he's by the tracks.

Lewis describes the phenomenon this way:

> The corpus [of the subject's beliefs] is fragmented. Something about the way it is stored, or something about the way it is used, keeps it from appearing all at once. It appears now as one consistent corpus, now as another. The disagreements between

> the fragments that appear are the inconsistencies of the corpus taken as a whole. (Lewis 1982: 436)

What we want, then, is a way of modeling Lewis's total doxastic state that allows us to represent him as having two distinct systems of belief, each of which drives some proper part of his total behavior.

Here is a natural and fairly conservative proposed revision of a unified model to capture this thought: rather than representing a subject's doxastic state with a set of worlds, or a credence function, or a set of propositions, we represent it with a set of sets of worlds, a set of credence functions, or a set of sets of propositions. In general, we move away from a unified model by representing subjects' total doxastic states with sets of whatever kind of thing we used to use—one of the old kinds of representors per fragment.

In Lewis's example, we represent his doxastic state with a pair of sets of worlds, one of which includes only worlds in which Nassau Street and the tracks run parallel and north–south, the other of which includes only worlds in which Nassau Street and the tracks run parallel and east–west. This captures the fact that Lewis's belief that Nassau Street and the railroad tracks run parallel is 'always on' (because this information is present in both fragments), but his belief that they run east–west guides only some of his behavior, while his belief that they run north–south guides other bits of it. The big set characterizes the atlas, and the members—each of which is an object of the kind that we were inclined to use to represent doxastic states when we were using a unified model—characterize the particular maps.

So we use one of the old-fashioned things to represent this sometime-active-in-behavior-guidance belief state, one of those things to represent that sometimes-active-in-behavior-guidance belief state, and the set containing both of those things to represent Lewis's total

doxastic state. We represent Lewis as doxastically fragmented—as having a representational system in which different doxastic representations, containing different information, are available for and active in behavior guidance in different circumstances (and with respect to different tasks).

To borrow some terminology from Yalcin (2018): let's use 'belief state' to talk about the fragments—the particular maps by which we steer some bit of behavior some of the time. (Each fragment, on the current proposal, is represented by one of the things the unified model uses to represent the totality of a subject's system of beliefs—for example, a set of possible worlds, a credence function, or a set of propositions.) And let's use 'doxastic state' (sometimes 'overall doxastic state' or 'total doxastic state') to talk about the whole atlas—the total system of belief, the total specification of how things are doxastically with an agent. (On the current proposal, represented by a set of whatever kinds of things we're using to represent belief states.)

One thing this does is let us keep a system where our individual belief states are consistent, probabilistically coherent, and logically closed, but our total system of beliefs isn't. (This was one of Lewis's, and also Stalnaker's, central motivations for going for a fragmented picture, and Lewis 1982 and Stalnaker 1991 are, I think, both plausibly interpreted as advocating for this sort of picture.) More importantly, it allows us to distinguish between the belief state which represents both the tracks and the street as running north–south, which guides some of Lewis's behavior, and the belief state which represents both as running east–west, and to find a home for both within our model of Lewis's total doxastic state.

It also allows us to capture (at least partially—see the next section) what's happening in cases of recognition and recall and of failure to bring to bear. We can represent Lucille's total doxastic state as containing the information that the youngest von Trapp child's name was

'Gretl.' The belief state that she brings to bear—that guides her behavior—in response to some stimuli does encode that information (by including only 'Gretl' worlds, or by including or assigning high probability to the 'Gretl' proposition). But not all of her belief states encode that information. In particular, the one that guides her responses to the *wh*-question doesn't. And both kinds of belief states can find a home in a single overall doxastic state (mutatis mutandis for failure-to-bring-to-bear cases like the power outage case—some of my belief states encode the information about power to the TV and power to the computer coming from the same source, but not all of them do).

That's progress. It gives us a way of representing subjects as having total systems of belief that aren't consistent, aren't closed under entailment, and which leaves room for only some of the information that's in the agent's total system of beliefs to be active in guiding any particular bit of behavior.

But it's not yet enough, for reasons we'll look at in the next section.

4. Complications and Research Questions: Constructing a Fragmented Framework

4.1. Specifying Spheres of Behavior Guidance

One reason why it's not yet enough is that it doesn't provide us with the resources to say when a subject steers by which map. For example, it doesn't allow us to distinguish Lewis, who steers by the map according to which both Nassau Street and the railroad tracks run north–south while

he's on the train, and by the map according to which both run east–west when he's on Nassau Street, from his counterpart who does the reverse.

To capture that sort of difference between believers—differences in the circumstances and domains of behavior in which they bring the various elements of a common repertoire of belief states to bear—we will need, at least, to supplement our models by adding something to our representations that identifies when our subject steers by which map. Representing a subject's doxastic state with a set of credence functions, for example, lets us identify the class of belief states that sometimes guide some bit of the subject's behavior. But it doesn't let us say anything about which bits of behavior are guided by which belief states. Within the atlas metaphor, this sort of model fully specifies what's on the pages of the atlas but doesn't say anything about which pages the subject looks at to guide which bits of behavior. It's reasonable to be dissatisfied with that limitation, because those sorts of differences between subjects are the sorts of things that we might want to be able to theorize about, and if we're going to theorize about them it would be convenient to have a theoretical framework that can represent them.

So here is one way in which we are likely to want to supplement our models: we add, in our representation of a subject's overall doxastic state, an element (or elements) that characterizes each particular belief state's behavior-guiding role. So now we might, for example, represent doxastic states with sets of pairs, one element of which is the kind of thing that we used to use to represent unified doxastic states (credence functions, sets of worlds, consistent and closed sets of propositions), and the other element of which has the role of characterizing the scope of the subject's behavior that's guided by that particular fragment. (The role of the first element of the pair is to characterize the map; the role of the second is to characterize when, and with respect to which bits of behavior, that map is the one by which the subject steers.) (The access tables in

Elga and Rayo, Chapter 1 in this volume, which pair elicitation conditions with specifications of the information available in that condition, are clearly an instance of this sort of view. Yalcin's 2018 model of question-sensitive belief looks like another.)

So, for example, we might represent subjects' total doxastic states with not just sets of credence functions, but sets of <credence function, context type> pairs. This would give us the resources to represent differences among subjects in terms of when particular bodies of information are brought to bear in the guidance of particular bits of behavior.

There is a choice point in the construction of our fragmented framework here, in what kind of object we use to identify the behavior-guiding role of a belief state. Some candidates include: questions, tasks, features of environments, circumstances, physical locations, incoming stimuli This is, as Dirk Kindermann (personal conversation) put it to me, 'the vexed question of how to individuate fragments.' The choices here are not totally straightforward and are likely to depend on our theoretical purposes and our views about the particular patterns of variation that we're going to want our theory to be able to capture. (See Bendana and Mandelbaum, Chapter 3 in this volume, Elga and Rayo, Chapter 1 in this volume, and Yalcin 2018 for some discussion and some options.)

One reason why the question of what to do at this choice point is not straightforward is that perhaps the simplest proposal—moving to sets of pairs of old-fashioned representors of belief states and specifications of contexts in which those belief states are active in behavior guidance—is only a first, and probably inadequate, step. It probably won't be enough just to carve up fragments/belief states by the times or types of circumstances in which they're active in behavior guidance.

4.2. Synchronic Fragmentation

That probably won't be enough because it's very plausible that we'll need to allow for synchronic fragmentation, in which, at a given time, one aspect of the agent's behavior is driven by one belief state while another aspect of the agent's behavior is driven by a different one.

Here's a version of a case from Stalnaker (1991: 439) (the shrewd but inarticulate chess player) to make the point:

Ron is a shrewd chess player—he consistently plays very well, and his expert play is best explained by attributing to him a rich system of sound beliefs about chess strategy. But Ron is terrible at articulating chess strategy, and he gives terrible advice. Sometimes, Ron sits in his house's common room and plays expert chess, all the while dispensing sincere but terrible chess-playing advice to his classmates, who are playing their own games.

We're going to want to appeal to one belief state to explain Ron's shrewd chess-playing, and a different one to explain his terrible advice, even if he's doing both at the same time. So what we'll want from a specification of the behavior-guiding role of a particular fragment or belief state will probably not be just a specification of a context for the agent to be in, such that in that kind of context that fragment drives all of the agent's (belief-governed) behavior.

We'll instead want to associate belief states with something more complicated that specifies which kinds of behavior they guide in which circumstances (Elga and Rayo, Chapter 1 in this volume, and Yalcin 2018 are responsive to this concern. Bendana and Mandelbaum's proposal in Chapter 3 of this volume seems perhaps susceptible to this sort of criticism).

So choosing just how to elaborate a fragmented model in a way that captures variation across believers—in terms of which fragments are active with respect to which circumstances or domains of behavior—is not straightforward.

4.3. Constraining the Fine-Grainedness of Spheres of Behavior Guidance to Avoid Triviality

At this point, it's worth flagging a general concern about fragmented models that will impose some constraints on what we say here: a concern about the danger of a fragmented framework's collapsing into triviality. Elga and Rayo raise the concern this way: 'One way for a theory to be explanatory is for it to identify patterns and show that relevant facts are instances of those patterns. Access tables [Elga and Rayo's fragmented models] are explanatory in at least this sense … This assumes, however, that the access table's elicitation conditions are not individuated too finely. Otherwise, an access table might become a mere listing of overly specific dispositions, and so fail to provide useful explanations of behavior' (Chapter 1 in this volume, p. xx). Unconstrained fragmented models of belief—on which elicitation/behavior-guidance conditions are unconstrainedly fine-grained and fine-tuned—threaten to just become unilluminating redescriptions of the subject's behavioral dispositions in needlessly baroque terms, or at least threaten to add nothing of theoretical interest beyond a specification of the subject's behavioral dispositions.

When we're allowed to vary which belief state guides different bits of a subject's behavior in different circumstances, constructing a pattern of beliefs to attribute that fits with observed behavior becomes too easy. We can find a fragmented doxastic state to attribute to any agent, and 'explain' any pattern of behavior, if we're free to vary the belief state that's driving their behavior in as fine-grained and fine-tuned a way as we like. But there should be *something* that a subject could do—some pattern of behavior they could display—that would make it inappropriate to interpret them, or some particular bit of their behavior, as driven by a system of beliefs (and desires). And there should be some principled limitations on how finely we cut the

behavior-guiding scope of a subject's belief states, so that our attributions of doxastic states to subjects are genuinely explanatory and theoretically interesting. So we will want to impose some extra constraints. Which constraints are attractive will depend on what we take our project to be.

Here we have another choice point. What constraints shall we impose on the proliferation of belief states in order to avoid collapse into triviality and maintain the status of doxastic states as explanatory of behavior rather than as a mere summary or redescription of it? One such constraint could be the requirement that there be some sort of meaningful correspondence to the subject's actual representational architecture (in Chapter 3 of this volume, Bendana and Mandelbaum go this way). Another would be to impose interpretationist pressure from other constraints on interpretation (Elga and Rayo, Chapter 1 in this volume, and Williams forthcoming go this way). I will not engage with these questions here and instead just want to flag this as an open (and urgent) research question for fragmentationists.

4.4. Updating Fragmented Doxastic States

So: suppose we've got an adequately constrained component of our fragmented model whose role is to represent each fragment's behavior-guiding scope. We are still not done building our fragmentationist framework for representing subjects' doxastic states. We will probably also want something in our model that indicates the pattern of responsiveness to evidence of the belief state that guides our subject's behavior in a given domain. (At least we'll want this on the assumption that we want to be able to model not only a subject's doxastic state at a particular time but also the ways in which the subject's overall doxastic state changes over time.)

On a fragmentationist model, it's unlikely that we will want to say that all of a subject's belief states are uniformly updated by all incoming evidence. And we will probably want our

models to have the resources to model variation in terms of which fragments of a subject's doxastic state are receptive (and perhaps also just how they are receptive) to which sorts of updates.

One obvious reason to resist a picture on which all of the belief states that make up a subject's overall doxastic state are uniformly updated in response to every incoming source of evidence is that this sort of uniform update will tend to wash out any fragmentation we started with, as the subject's belief states grow more and more similar as they incorporate more and more of the same information. Suppose I start off in a doxastic state containing a fragment that includes the information that the youngest von Trapp child's name is 'Gretl' and another fragment that doesn't. Then I watch the movie, note that the youngest von Trapp child's name is 'Gretl,' and update my doxastic state. If all of the belief states that make up my overall doxastic state are updated uniformly, I should then wind up in an overall doxastic state according to which the information that the youngest von Trapp child's name is 'Gretl' is always available. But this is not (or at least not uniformly) what's likely to happen, and so the assumption of uniform updating is not going to give us a fragmented model that's well positioned to account for phenomena like the recognition/recall distinction.

Uniform updating would also make it mysterious how we come to be fragmented in the first place. Presumably, the reason why I'm now (or was before I started writing this chapter) fragmented about the information that the youngest von Trapp child's name is 'Gretl' is that I learned it—I updated with it on the basis of some evidence—but that update left me in a state in which that information was not uniformly available for behavior guidance (a fact we are modeling by attributing to me a doxastic state that includes a fragment that doesn't include the

'Gretl' information). So there must be some fragment that didn't get updated with that information.

In general, one of the phenomena that we will probably want to use our fragmented models to theorize about is the following sort of situation: some new evidence comes in, and that information becomes available to the subject for behavior guidance for some purposes and in some circumstances but does not become available for behavior guidance for all purposes and in all circumstances. The way this will be reflected in our models is by encounters with evidence frequently resulting in updates to some, but not all, of the belief states that make up the subject's overall doxastic state (see Bendana and Mandelbaum, Chapter 3 in this volume, for discussion of this sort of phenomenon as a motivation for redundant representation).

Another central fragmentation-motivating phenomenon is the persistence of beliefs despite receiving evidence that should, on a unified model, eliminate them (Anderson et al. 1980; Anderson and Sechler 1986; Slusher and Anderson 1989) and the failure of the consequences of things we update with to 'percolate through' our total system of belief. (One species of closure failure like this occurs when we fail to draw certain consequences from new beliefs—we believe that if *P* then *Q*, we update with *P*, but we don't form a belief that *Q*. The power outage case from Section 2 is an example.)

Sometimes when I update by adding *P*, I don't update with all of the consequences of *P*. An attractive way to model this is to say that not every fragment is going to be sensitive to every update. (So when I believe that if *P* then *Q* and update with *P* but don't wind up believing *Q*, that's because the fragment that contained my *if P then Q* belief didn't get updated with *P*.) When the information that *P* comes in, we often update some, but not all, of the maps in our atlas with *P* and its consequences. (At least one version of the power outage case is like this—I

updated my no-power-then-no-TV belief state with the evidence about the power outage, but I didn't update the belief state where my no-power-then-no-computer belief lives.)

Here is another instance of the same kind of phenomenon: sometimes I update with *P* and come to believe both *P* and a bunch of consequences that follow from *P* plus other things I believe. Then my evidence for *P* is later discredited, and I extract *P*. But I won't necessarily extract all of the downstream conclusions that I relied on *P* to draw (see Bendana and Mandelbaum, Chapter 3 in this volume).

If we want our models to be well suited to representing and theorizing about this sort of phenomenon, we will want another element in our model (in addition to a specification of the informational content of the various fragments and a specification of the domains in which each fragment is active in behavior guidance), whose role is to specify what kinds of evidence or inputs each particular belief state is sensitive to.

5. Complications and Research Questions: Normative Questions about Fragmented Belief States

I hope to have made a plausible case that the project of constructing fragmented models for theorizing about belief is well motivated, but also not straightforward. I now want to draw attention to some normative questions that arise once we've moved to a fragmented picture of belief.[4]

[4] I will also drop a footnote briefly discussing a couple of metaphysical questions that arise.

The questions about the motivation for and construction of fragmented frameworks have received some attention in the literature, and the amount of attention they've been receiving has been ramping up in recent years. The normative questions I'm about to discuss have received very little attention to my knowledge,[5] but they are important questions for fragmentationists to take up. Specifically, fragmentationists will need to address questions about whether and under what conditions it is rational to be fragmented (Section 5.1); about the comparative rationality of differently fragmented doxastic states (Section 5.2); about the rationality of transitions between fragmented doxastic states (Section 5.3); and about the role of fragmented doxastic states in the rationalization of behavior (Section 5.4).

My goal in this section will be to put these questions clearly on the table and to make a case for their urgency. I won't do much in the way of offering answers to them, both because I do not feel confident about how to answer most of them and because that would be another (much longer!) paper.

5.1. First Normative Question: The Rationality of Fragmentation

Is it always rationally better, cognitive capacities permitting, to be unified than fragmented? It's tempting to say, 'Yes, of course.'

But this turns out not to be so straightforward. I've argued (I think convincingly; Egan 2008) that fragmentation can serve as a useful damage-control device in cases where agents have belief-forming mechanisms that are liable to go wrong. The fact that agents have such fallible

[5] They have received *some* attention: see Cherniak (1983, 1986), Borgoni (Chapter 5 in this volume), Egan (2008), Greco (2014a), Johnson (2020), and Yalcin (Chapter 6 in this volume).

belief-forming mechanisms can make it more rational, in certain kinds of cases, to be fragmented than to be unified. (The idea is that, if you're stuck forming beliefs in an unreliable way, it's useful to be able to quarantine those unreliably formed beliefs so that you can keep them from infecting your whole system of beliefs by way of free-wheeling inference, and so that you can keep them from misdirecting too much of your behavior by restricting their behavior-guiding role.)

Another kind of case of potential rational fragmentation is that of practically indispensable but theoretically dubious beliefs. Suppose, for example, that you take a philosophy class and you become convinced that there's no free will. Or that we shouldn't form beliefs on the basis of induction. Or that there aren't any reasons. Or that there's no causation.

But you can't hold any of those beliefs in mind and still function in the world, because they undermine practical deliberation so badly.

Maybe, in these kinds of cases, fragmentation is a rational response: you don't discard the philosophical beliefs, and you don't discard the practically indispensable beliefs that they undermine, either. Instead, you fragment; you allow your seminar room argumentation to be driven by the philosophical beliefs and your post-seminar billiard-playing to be driven by your ordinary beliefs.

So it's at least not straightforward that it's always better, rationally speaking, to be unified than fragmented (see also Yalcin, Chapter 6 in this volume, for further discussion of this issue). Certainly it's not the case that, for any individual *a*, evidential state *e*, and pair of doxastic states *x* and *y*, if *x* is unified and *y* is fragmented, it's more rational for *a* to be in *x* than *y*, given *e*. (An easy case: the evidence in *e* strongly supports *P*, unified state *x* includes a belief that not-*P*, and fragmented state *y* is split on *P* and suspension. The fragmented state is clearly better because it

is in a clear sense more evidence-responsive. The troublemaking cases in Egan 2008 have a similar sort of structure. They're cases in which, because of the imperfections in the subject's belief-forming mechanisms, the unified state that the subject would be in if they were unified is one with (unified) commitments that are badly supported by the subject's evidence, while the available fragmented state is one that contains a fragment with commitments that are badly supported by the evidence, but also a fragment with commitments that are well supported by the evidence.)

5.2. Second Normative Question: What Should We Say about the (Absolute and Comparative) Rationality of Different Fragmented Doxastic States?

A more general question also presents itself: What should, and what can, we say about which fragmented doxastic states are rational and which are irrational? And about which fragmented states are more or less rational than others? The answers here also seem to be less than straightforward.

Similar kinds of cases to those just discussed will undermine (I think) some straightforward answers, for example that *less* fragmented states are always rationally better. If one's evidential state supports *P*, it seems plausible that it's better to be in a doxastic state that somewhere encodes *P* than one that does not encode it at all, even if that requires additional fragmentation.

It also seems better, if one's evidential state supports *P*, to be in a doxastic state where *P* is available for more purposes rather than fewer.[6] It seems likely that we will want to be able to say something about comparisons between fragmented doxastic states with respect to how broadly available certain information is (and how broadly deployed for behavior guidance certain *mis*information is). We obviously won't be able to do this in a satisfactory way by, for example, counting how many fragments the proposition appears in. And it's not straightforward how to measure the relative sizes of spheres of behavior guidance.

In addition, we may want to take into account the extent to which one's pattern of information access is *properly targeted*. There's no need for me to bring to bear all of my beliefs about the layout of Baltimore when navigating New Brunswick. And there's no need for me to

[6] Carolina Flores (personal conversation) points out that the philosophy cases above seem like a potential counterexample: maybe some philosophical skeptical beliefs are quite strongly supported by my evidence, but it is, at least in terms of *practical* rationality, better if they have a sharply circumscribed behavior-guiding role. Relatedly, we can think about cases in which the costs of acting on the basis of a false belief that *P* and those of acting on the basis of a false belief that not-*P* aren't symmetrical. Sometimes, it's better all things considered to be guided by a belief that there's a predator nearby even if the evidence better supports believing that there isn't, because the costs of guiding behavior by a false predator-present belief are much lower than the costs of guiding behavior by a false predator-absent belief (see for example Sterelny 2003). Also relatedly, maybe it's best all things considered for a great deal of my behavior to be guided by an assessment of my own abilities that's overly optimistic, given my evidence (see for example Elga 2005; Bendana and Mandelbaum, Chapter 3 in this volume; Taylor and Brown 1988, 1994). So things are probably more complicated than I'm letting on in the main text.

bring to bear all of my beliefs about New Brunswick when I'm driving around Baltimore. Given limited cognitive resources and processing power, I'm doing pretty well if I'm fragmented in such a way that my Baltimore beliefs, but not my New Brunswick ones, are available for behavior guidance while I'm navigating Baltimore, and my New Brunswick beliefs, but not my Baltimore ones, are available for behavior guidance while I'm navigating New Brunswick. But I'm doing very badly if I have the reverse pattern of access (see Cherniak 1983: sec. 5 for discussion of this sort of issue).

Now it's not completely clear whether we want to count this kind of failure of proper targeting—in which I fail to bring the right information to bear in the guidance of the bits of behavior where it's most likely to be helpful—as a *rational* failing. But it's not completely clear that we don't. If it is a rational failing, it would be nice to be able to state the relevant principle(s) for the rational ordering of fragmented doxastic states, or of better- and worse-making features of them.

And whether it's a rational failing or a failing of some other kind, there is a market for a story about this—there's a kind of criticism and evaluation of fragmented states that we'd like to be able to offer on grounds of the information's being available for the guidance of not enough, or the wrong kinds of, behavior. Some fragmented states do better and others do worse in this regard. But it's not obvious just how to spell out the general principles here.

Another thing that's not obvious is whether the type of rational evaluation at issue in this section (and the following) is *practical* or *epistemic*. Drawing that distinction between kinds of rational criticism and praise of particular patterns of fragmentation also seems like something we'd like to be able to do, and it's not clear that the answers will fall directly out of the way we draw the distinction between ways of criticizing and praising unified states.

5.3. Third Normative Question: What Should We Say about the (Absolute and Comparative) Rationality of Different Transitions between Fragmented States?

Here is a place where there is quite a clear price to pay for moving from a unified to a fragmented model. We have made a lot of progress in thinking about rational transitions between unified states—there is an enormous literature on both the rational updating of credence functions and the rational updating of unified bodies of binary on/off belief (see Chignell 2018; Easwaran 2011a, 2011b; and Huber 2016 for surveys).

But it's not at all obvious what we ought to say about rational transitions between fragmented states, and it's not clear that much straightforwardly follows from what we know about unified states.

One reason to think that we won't just be able to read off our principles for evaluating transitions between fragmented states from already well-worked-out principles for evaluating transitions between unified states is that updates of fragmented states in response to evidence aren't in fact, and perhaps ought not always to be, uniform. (So it's not, for example, that when evidence *P* comes in, a subject uniformly updates every credence function in their overall doxastic state by conditionalizing on *P*. This certainly isn't what actually happens, and it's also not clear that it's always what *ought* to happen—see Sections 4.4 and 5.1.)

Another reason is that some of the transitions between fragmented states that we'll want to be able to theorize about are fragmentation-specific—they aren't kinds of transitions that will appear in a unified picture, and so our theorizing about unified pictures is unlikely to provide us with much guidance about what to say about them. One such transition is the kind that happens

when we recognize an inconsistency in our beliefs and set out to resolve it—the process of *defragmentation*. Some responses to the discovery that we harbor inconsistent beliefs seem better than others. (Endorsing explosion, and thus inferring everything, is very bad; going in for some sort of process of assessing the evidential basis for the two inconsistent beliefs and excising the less-well-supported one seems pretty good. And some versions of the assessment-and-excision process are better than others.) It would be nice to have some general and systematic things to say about this, if such things are available. And if there aren't any systematic generalizations available, that would also be interesting, and it would be interesting to know why that's so.

Relatedly, there are questions about relations of evidential support between bodies of evidence and doxastic states which look as if they won't be straightforward to answer. There is a great deal of work that's been done exploring the relations of evidential support between bodies of evidence and unified states and characterizing which unified states are better supported by, and more rational in light of, which bodies of evidence. The analogous questions about fragmented states, and which fragmented states are better supported by which bodies of evidence, have not been well explored.

5.4. Fourth Normative Question: What Should We Say about the Role of Fragmented States in Rationalizing Behavior?

Again, we have a lot of work on the role of *unified* states in rationalizing behavior to draw on. But it is not so clear how to carry that over to fragmented states. Standard forms of decision theory all make use of unified models of subjects' beliefs (and desires—but that is a different can of worms). Is there anything to say about the role of fragmented states in rationalizing behavior that's similarly systematic? And if so, what does it look like? These questions are thus far largely

unexplored (although Elga and Rayo (unpublished manuscript) have made helpful progress on this front).[7]

6. Conclusion

[7] There are also a couple of metaphysical questions that arise which I'm putting in a footnote because I'm not confident about my presentation of them, though I think they are worth drawing attention to. First, there's the question of what a fragmented metaphysics of belief should look like. Accommodating fragmentation imposes constraints on one's metaphysics of belief—in particular, if we're going to be fragmentationists, we had best not formulate the functional role of belief (if we're functionalists) or our principles of doxastic interpretation (if we're interpretationists) in a way that presupposes unification. We'll want to make sure that fragmented believers, and fragmented belief states, count as believers and belief states. (And we'll want to make sure that subjects who aren't properly interpretable as unified believers, or who don't have any states that play the functional role of always-on, unified-model beliefs, still count as having beliefs.) So we'll want to avoid, for example, building our functionalist or interpretationist accounts on the foundation of a decision theory that assumes a unified model. Second, there's the question of how much, and what kind of, fragmentation is compatible with thinking that there's a single thinking, acting subject that we're modeling (rather than no enduring subject at all, or more than one). What kind of pattern of imperfect, variable information access, if any, would warrant saying that there's no persisting agent there, or that there's more than one agent, controlling a common body in turns? One can imagine some clear science-fictional cases in which it seems plausible to say that there's no persisting agent, or that there's one who's present during the day and another at night. But exactly where the boundaries will be doesn't seem to be an easy question. (Thanks to an anonymous reviewer for raising it.)

The fragmentationist project is, it seems to me, well motivated. There are clear limitations to the capacity of unified models of belief to represent, and facilitate theorizing about, some ubiquitous and consequential doxastic phenomena. There's a market for a framework that allows us to represent and theorize about the ways in which our systems of belief are disunified, such that information that's available to a subject for one purpose isn't always available for another.

We have not yet settled just what such models should look like. I've drawn attention to a few issues and questions that arise in the project of constructing them. We have, in addition, only scratched the surface of questions about the rational evaluation of fragmented doxastic states and transitions between them, as well as questions about the relation between fragmented doxastic states and rational action.

I should also note that, in addition to raising tricky questions in need of attention, fragmented models hold the promise of constructive application to existing problems. So far, we have seen applications of the tools of fragmented models to the problem of old evidence (Fleisher unpublished manuscript), the problem of logical omniscience (an unsurprisingly large literature here, but for example Stalnaker 1984, 1991; Elga and Rayo, Chapter 1 in this volume; Rayo 2013; and Yalcin 2018. On implicit bias, see Huebner 2009 and Bendana and Mandelbaum, Chapter 3 in this volume; on higher-order evidence, see Greco 2019; on the preface paradox, see Cherniak 1986; and on the KK principle, see Greco 2014b, to name a few).

My goals in this chapter have been modest—this chapter is for the most part a plea for help rather than an argument for a particular view. What I hope to have done is to motivate the fragmentationist project by setting out some of the most compelling phenomena that fragmented models are meant to be responsive to and by drawing attention to unresolved research questions

that arise once one decides to go fragmentationist. I will be delighted if progress on these questions quickly renders this discussion of them hopelessly outdated.[8]

References

Anderson, C., and Sechler, E. (1986), 'Effects of Explanation and Counterexplanation on the Development and Use of Social Theories', Journal of Personality and Social Psychology 50/1: 24–34.

Anderson, C., Lepper, M., and Ross, L. (1980), 'Perseverance of Social Theories: The Role of Explanation in the Persistence of Discredited Information', Journal of Personality and Social Psychology 39/6: 1037–1049.

Braddon-Mitchell, D., and Jackson, F. (2007), The Philosophy of Mind and Cognition (2nd edn, Blackwell).

Cherniak, C. (1983), 'Rationality and the Structure of Human Memory', Synthese 57/2: 163–186.

Cherniak, C. (1986), Minimal Rationality (MIT Press).

Chignell, A. (2018), 'The Ethics of Belief', Stanford Encyclopedia of Philosophy (Spring 2018 edition). URL: https://plato.stanford.edu/archives/spr2018/entries/ethics-belief.

Easwaran, K. (2011a), 'Bayesianism I: Introduction and Arguments in Favor', Philosophy Compass 6: 312–320.

Easwaran, K. (2011b), 'Bayesianism II: Criticisms and Applications', Philosophy Compass 6: 321–332.

Egan, A. (2008), 'Seeing and Believing: Perception, Belief Formation and the Divided Mind', Philosophical Studies 140/1: 47–63.

Elga, A. (2005), 'On Overrating Oneself… and Knowing It', Philosophical Studies 123: 115–124.

Elga, A., and Rayo, A. (unpublished manuscript), 'Fragmented Decision Theory'.

Fagin, R., Halpern, J., and Vardi, M. (1995), 'A Nonstandard Approach to the Logical Omniscience Problem', Artificial Intelligence 79/2: 203–240.

Field, H. (1986a), 'Critical Notice: Stalnaker, Robert Inquiry', Philosophy of Science 53/3: 425–448.

Field, H. (1986b), 'Stalnaker on Intentionality: On Robert Stalnaker's Inquiry', Pacific Philosophical Quarterly 67: 98–112.

Fleisher, W. (unpublished manuscript), 'Fragmentation and Old Evidence'.

---

[8] Thanks to Austin Baker, Will Fleisher, Carolina Flores, Anton Johnson, Dirk Kindermann, Eric Mandelbaum, a reviewer for this volume, an audience at Johns Hopkins University, the participants in a joint seminar on fragmentation co-taught by me, Adam Elga, and Agustín Rayo, and especially to Adam Elga and Agustín Rayo themselves for comments and suggestions on this chapter and for conversations that have shaped my understanding of these issues.

Greco, D. (2014a), ‘A Puzzle about Epistemic Akrasia’, Philosophical Studies 167: 201–219.

Greco, D. (2014b), ‘Iteration and Fragmentation’, Philosophy and Phenomenological Research 88/1: 656–673.

Greco, D. (2019), ‘Fragmentation and Higher-Order Evidence’, in M. Skipper and A. Steglich-Peterson (eds.), Higher-Order Evidence: New Essays (Oxford University Press), 84–104.

Huber, F. (2016), ‘Formal Representations of Belief’, Stanford Encyclopedia of Philosophy (Spring 2016 edition). URL: https://plato.stanford.edu/archives/spr2016/entries/formal-belief.

Huebner, B. (2009), ‘Trouble with Stereotypes for Spinozan Minds’, Philosophy of the Social Sciences 39/1: 63–92.

Johnson, D. (2020), ‘Deliberations on Deliberation’, PhD thesis, Rutgers University.

Lewis, D. (1979), ‘Attitudes De Dicto and De Se’, Philosophical Review 88: 513–543.

Lewis, D. (1982), ‘Logic for Equivocators’, Noûs 16/3: 431–441.

Lewis, D. (1986), On the Plurality of Worlds (Wiley-Blackwell).

Ramsey, F. (1931), The Foundations of Mathematics (Kegan Paul).

Rayo, A. (2013), The Construction of Logical Space (Oxford University Press).

Schwitzgebel, E. (2001), ‘In-Between Believing’, Philosophical Quarterly 51: 76–82.

Schwitzgebel, E. (2002), ‘A Phenomenal, Dispositional Account of Belief’, Noûs 36/2: 249–275.

Slusher, M., and Anderson, C. (1989), ‘Belief Perseverance and Self-Defeating Behavior’, in R. Curtis (ed.), Self-Defeating Behaviors: Experimental Research, Clinical Impressions, and Practical Implications (Plenum Press), 11–40.

Soames, S. (1985), ‘Lost Innocence’, Linguistics and Philosophy 8/1: 59–71.

Soames, S. (1987), ‘Direct Reference, Propositional Attitudes and Semantic Content’, Philosophical Topics 15/1: 47–87.

Speaks, J. (2006), ‘Is Mental Content Prior to Linguistic Meaning? Stalnaker on Intentionality’, Noûs 40/3: 428–467.

Stalnaker, R. (1984), Inquiry (MIT Press).

Stalnaker, R. (1991), ‘The Problem of Logical Omniscience, I’, Synthese 89/3: 425–440.

Sterelny, K. (2003), Thought in a Hostile World (Blackwell).

Taylor, S., and Brown, J. (1988), ‘Illusion and Well-Being: A Social Psychological Perspective on Mental Health’, Psychological Bulletin 103/2: 193–210.

Taylor, S., and Brown, J. (1994), ‘Positive Illusions and Well-Being Revisited: Separating Fact from Fiction’, Psychological Bulletin 116/1: 21–27.

Williams, J. (forthcoming), ‘Commitment Issues in the Naïve Theory of Belief’, in A. Egan, P. van Elswyk, and D. Kindermann (eds.), Unstructured Content.

Yalcin, S. (2018), ‘Belief as Question-Sensitive’, Philosophy and Phenomenological Research 97/1: 23–47.