

UCLA

UCLA Electronic Theses and Dissertations

Title

Contradictions and the structure of reasoning

Permalink

<https://escholarship.org/uc/item/8sn3v8pp>

Author

Johnston, Kimberly

Publication Date

2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Contradictions and the structure of reasoning

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Philosophy

by

Kimberly Johnston

2022

© Copyright by
Kimberly Johnston
2022

ABSTRACT OF THE DISSERTATION

Contradictions and the structure of reasoning

by

Kimberly Johnston

Doctor of Philosophy in Philosophy

University of California, Los Angeles, 2022

Professor Sherrilyn Roush, Co-Chair

Professor Joshua David Armstrong, Co-Chair

In the dissertation I explore the role of inconsistency in human reasoning as a way into broader questions about the structure of reasoning processes. In each of the next three chapters I defend the idea that certain types of inconsistency can be rational against a different pressure. These defenses fill out a picture of reasoning on which consistency plays an important role in managing specific modes of reasoning and in how we justify our beliefs. But inconsistency can arise in a well-functioning system, since it is in general better to draw on an inconsistent belief state than to force the system as a whole to be consistent or to only draw on a consistent subset of the total belief state in any given reasoning situation.

We can helpfully divide discomfort with inconsistent belief states into different types: worries about how an agent could come to be in an inconsistent belief states and worries about how an agent could (or should) function once they are in an inconsistent belief state. The latter worry is addressed in chapter 2, where I address strengths and weaknesses in the fragmentation view. The former worry can be further divided into concerns about how an agent could end up in an inconsistent belief state full stop and how they could do so rationally.

I discuss how a rational agent could end up in an inconsistent belief set in chapter 3, defending myself from a possible objection from Niko Kolodny (using a view of rationality as correct response to different pressures) along the way. I argue for the existence of inconsistent beliefs full stop in chapter 4; inconsistent beliefs are possible on an interpretationist framework despite some Davidsonian worries to the contrary. In chapter 5 I pause to explore what these results mean for our understanding of human reasoning, including: the value of consistency, integrating different methods of reasoning, social reconstructions of reasoning, and how we should be evaluating agents.

The dissertation of Kimberly Johnston is approved.

Alexander Jacob Julius

Gabriel Jae Greenberg

Calvin G. Normore

Joshua David Armstrong, Committee Co-Chair

Sherrilyn Roush, Committee Co-Chair

University of California, Los Angeles

2022

For Ian

TABLE OF CONTENTS

1	Introduction	1
1.1	Inconsistent beliefs	1
1.2	Notes on methodology	6
2	Isolating inconsistencies	12
2.1	Fragmentation	13
2.2	Against fragmentation	20
2.2.1	Norby against fragmentation	20
2.2.2	Fragments and belief subsystems	22
2.2.3	Resolving	25
2.3	Conclusion	31
3	Incomparability and inconsistency	36
3.1	Consistency and evidential norms	38
3.2	Rational pressures and conflicts	42
3.2.1	Fogal on primacy of pressure	43
3.2.2	Pryor on weighing norms	44
3.2.3	Worsnip and Kolodny on inter-level consistency	46
3.3	Comparability and the threshold for belief	50
3.3.1	Kolodny's reliance on comparability	50
3.3.2	Etiology and belief formation	53
3.4	Conclusion	61

4	Interpreting the inconsistent	64
4.1	Davidson on inconsistent beliefs	66
4.2	Contexts of belief attribution	74
4.3	Examples of belief attribution	80
4.4	Attributing inconsistencies	85
4.4.1	Rationality again	85
4.4.2	Belief attribution again	89
4.5	Conclusion	92
4.5.1	Objections and replies	92
4.5.2	Concluding remarks	94
5	Implications	96
5.1	The structure of reasoning	98
	Bibliography	101

ACKNOWLEDGMENTS

I am extremely grateful to my committee. Josh Armstrong took me on as a student without hesitation when I was switching areas and trying to figure out what my philosophical project was. His advice – philosophical, personal, and professional – has been invaluable to me over the years. Sherri Roush is extremely good at pressing me on my examples, and I am a better philosopher for it. She also kept me on track many times when I tended to veer off it. Calvin Normore has always challenged me to shift my thinking and suggested new ways of approaching difficulties I was confronted with. Gabe Greenberg is exceptionally good at asking specific and helpful questions and has helped me to improve my writing over the years. A. J. Julius helped me with this project before it was this project and led to me to consider things in a different light.

I also got a lot of help from other faculty members along the way, at UCLA and beyond. Thank you to Lije Millgram and Anya Plutynski for getting me into this mess in the first place. I got help I did not even know I needed from Sam Cumming, Barbara Herman, and Seana Shiffrin. Thanks are also due to Tyler Burge, Katie Elliott, Pamela Hieronymi, Andrew Hsu, David Kaplan, Jonah Schupbach, Shel Smith, Mariam Thalos, Sean Walsh, Laurel Westrup, and many other amazing faculty members at UCLA.

There are many other wonderful people I need to acknowledge, starting with Greg Antill, Lindsey Chambers, Taylor Doran, Ashley Feinsinger, Jonathan Gingerich, Arielle Hill-Moses, Will Reckner, and Ren Schaeffer for showing me the ropes in one way or another. I owe Milo Crimi, Christian De Leon, Amber Kavka-Warren, Bill Kowalsky, and Allison McCann for their support, especially these last few years. Thanks are due most of all to Ian Boon. Without him I would never have made it through. I would also like to express my gratitude to Rima Basu, Tristen Cardwell, Gabe Dupre, Mariko Green, Gabby Johnson, Tanya Kostochka, Kevin Lande, Ayana Samuel, Fémi Táíwò, Aaron West, and the graduate students at UCLA who help to make it such a fantastic place to do philosophy.

Finally, thanks are due to my mother for keeping me stubborn and teaching me to read voraciously, to my father for keeping me patient and teaching me to think carefully, and to my sister for keeping me alive and teaching me to see compassionately.

VITA

- 2010 B.A. Mathematics, University of Utah
- 2014 M.S. Philosophy, University of Utah
- 2016 M.A. Philosophy, UCLA
- 2017 C.Phil Philosophy, UCLA

CHAPTER 1

Introduction

1.1 Inconsistent beliefs

It is an often-unspoken assumption in many areas of philosophy that consistency is necessary for rationality. Here I will briefly review this standard view about the relationship between consistency and rationality before discussing some increasingly challenging examples for this view. While I think it is worth noting that there are parallel attitudes in the literature towards belief-like states and desire-like states, I will be focusing on belief-like states in this dissertation. For this reason, I will not be discussing arguments for the standard position regarding desire-like states and the examples are focused on inconsistent beliefs and not other incoherences (like incoherent intentions or intransitive preferences).

In most of the metaethics and practical reasoning literature, inconsistency or incoherence is by itself and automatically a condemnation whenever it is true of an agent or a set of intentions, desires or beliefs. Having an incoherent plan or incoherent intentions (having a set of intentions that cannot all be satisfied, like planning on being in two places at once) means that you are failing as an agent. And while desires or ends cannot be inconsistent in any straightforward sense, there is a related attitude towards conflicting desires or ends. When they are spoken of, it is assumed to be a bad thing that people have ends that cannot all be satisfied, and so conflict with one another. In attempts to formalize desire-like states, we always see constraints on what desires have to look like for the theory to work. In decision theory, for example, desires must be weakly ordered. This means that preferences must obey

transitivity. It also means that for any two ends or desires the agent must prefer one to the other or be indifferent between them. While constraints on preference orderings are not exactly consistency constraints, people who violate the transitivity constraint are often spoken of as having inconsistent preferences, so this seems to be related to the same intuitive notion of consistency or coherence being a good thing for agents.

When it comes to beliefs, a similar situation appears. The view that having inconsistent beliefs is obviously irrational is very prevalent. One source of this belief is the idea that classical logic has some important connection to reasoning. For example, it is common to find reference to classical fallacies (which are commonly explained using either classical logic or probability theory) in explanations of why some bit of reasoning is bad. While this relationship between classical logic and reasoning is often undeveloped, Lance Rips (1994) argues that deductive reasoning in the form of classical logical inferences is at the heart of thinking. Any such view will not be able to handle rational inconsistency. In classical logic, an inconsistent set of propositions is one that does not have a model. This has a couple of important implications for rationality. First, this means that the propositions in such a set cannot all be true. So, *prima facie*, since having false beliefs is epistemically irrational, it is irrational to have inconsistent beliefs. Second, it also means that one can prove anything taking that set as your set of premises. (This property, that anything follows from a contradiction in a particular logic, is called “explosion.”) So, if your beliefs are inconsistent, there is not much you can do with them within classical logic.

But there are a range of situations that make it seem as though people having inconsistent beliefs is extremely common, and some that motivate that it is rational. The existence of inconsistent beliefs is not by itself reason to think that it must be rational to be inconsistent. But if they are ubiquitous then it is a reason to re-think our relationship to our rationality standards: If everyone is irrational, then we need to get clear on a graded view of rationality and possibly reconsider if the standard is too harsh. I will now discuss several types of examples in turn before summarizing my plan for the rest of the dissertation.

First, it is important to be clear that it is certainly not the case that we should think that someone has inconsistent beliefs whenever they assert inconsistent things. The simplest case of this is belief change. If someone at one time asserts that lichen is a type of moss and later on asserts that lichen is not a moss, we can assume that they learned some new information and changed their mind in that time. We have no reason to interpret them as being synchronically inconsistent.

Another type of case where we certainly do not want to say that the person has inconsistent beliefs is cases where we have reason to think that the person is simply lying, or otherwise saying something that they do not believe. If my friend has an irrational fear of fungus and I want to get them to go hiking with me, I might lie and say that lichen is a type of moss. Even when paired with me saying to another person that lichen is not moss, that is not good evidence that I have inconsistent beliefs.

There are also cases where it seems like people do have inconsistent beliefs, but they are clearly irrational. Donald Davidson (2004a, 209) gives the following case: Carlos is about to take a driving test. His evidence supports the proposition that he will fail the test, and he is aware of this fact. He has already failed the test twice and gets a lot of discouraging feedback from his driving instructor. So, he (rationally) believes that he will fail the test. But he also believes that he will not fail the test. This is because the thought that he will fail is painful to him, especially as he tries to practice his driving skills. This pain, according to Davidson, causes a split in his beliefs and he ends up having both of the contradictory pair. There are many other examples in this vein, since people often have evidence to support a conclusion that it would be in some practical sense destructive to believe. (And so, can form beliefs that are practically beneficial but epistemically irrational.)

Sometimes the inconsistent belief set is the result of a much simpler mistake. Tamar Gendler gives a real-life example of her forgetting her wallet at an APA meeting. She asked a friend to borrow money because she had forgotten her wallet and then, money in hand, looked for her wallet to put the money in (Gendler 2008). This looks like a case where she did

have inconsistent beliefs, but the explanation is clear. She had a moment for forgetfulness and once the situation was made clear, simply updated her beliefs to be correct.

Some more interesting cases involve different points of view on the world or different ways of reasoning. One interesting example involves conflicts between a scientific or broadly theoretical attitude towards the world and a common sense and day-to-day attitude. Wilfrid Sellars (1962) referred to these as the manifest and the scientific image. These two attitudes towards the world can lead to inconsistent beliefs about the world around us. We think of objects as solid but also as made up mostly of empty space, for example. Both of these attitudes seem necessary for your reasoning about these objects, especially for people involved in the sciences. Another interesting conflict is that between on-the-ground reasoning about particulars and higher level reasoning about the reliability of that on-the-ground reasoning. This conflict leads to interesting cases like the paradox of the preface.

I will introduce one more case, due to Berislav Marušić (2015). Take an agent who has resolved not to smoke for the fourth time. It is complicated to say whether they should have the belief ⟨I will smoke again⟩ or the belief ⟨I will not smoke again⟩. They have reasonable observational evidence that they will fail again, but they also have evidence of a different kind that they will succeed. They are in a particular position to make beliefs about their own behavior true or false. This is structurally similar to a family of cases involving someone looking at themselves in a third personal way (i.e., as an arbitrary human with a reasonably predictable mental makeup and habits) but also having an attitude towards themselves as someone who can control their own behavior. Agents can think of themselves as a person like any other, but they also look at themselves in a particular way because they are themselves. These types of cases are further complicated by practical considerations, since it is dangerous to abdicate responsibility for one's own actions and view oneself in a purely third-personal way, but it can also be helpful to play tricks on yourself to aid in your chosen path of action. If I am only thinking about myself in a disinterested way, I might be able to effectively manage myself, but I run the risk of losing my sense of agency. But, if I fail to take the

fact that I am a fallible human into account, I risk being overconfident, not using all of the resources at my disposal, and failing.

These cases give us reason to re-think how consistency functions in our theories of reasoning. But, of course, I do not want to say that holding inconsistent beliefs is universally fine or that there is no merit to consistency in any situation. One of the things I want to determine in the course of this project is what role consistency plays in reasoning and why it is valuable. I do not want to discard the idea that it can be helpful to point out an inconsistency; I also do not want to discard the idea that, depending on the context, something being inconsistent is a problem. For example, inconsistency does, at least some of the time, seem to demand to be resolved before the agent can move on with reasoning about an issue or making a practical decision. So, of course, I do not think that all inconsistent beliefs should be maintained. Many inconsistencies should be ignored, as they are not worth resolving and may not be very useful in future reasoning processes. For example, if I accidentally form some inconsistent beliefs about proper welding technique while listening to a lecture, but I will never think about welding ever again, it is not worth my cognitive resources to sort that out. Some should be resolved by some deadline (where by resolve we mean that the agent believes one and stops believing the other). I will return to the question of which inconsistencies call for which treatment and why in the conclusion. Now I will turn to an outline of the rest of the dissertation.

In the dissertation I explore the role of inconsistency in human reasoning as a way into broader questions about the structure of reasoning processes. In each of the next three chapters I defend the idea that certain types of inconsistency can be rational against a different pressure. These defenses fill out a picture of reasoning on which consistency plays an important role in managing specific modes of reasoning and in how we justify our beliefs. But inconsistency can arise in a well-functioning system, since it is sometimes better to draw on an inconsistent belief state than to force the system as a whole to be consistent or to only draw on a consistent subset of the total belief state in any given reasoning situation.

We can helpfully divide discomfort with inconsistent belief states into different types: worries about how an agent could come to be in an inconsistent belief states and worries about how an agent could (or should) function once they are in an inconsistent belief state. The latter worry is addressed in chapter 2, where I address strengths and weaknesses in the fragmentation view. The former worry can be further divided into concerns about how an agent could end up in an inconsistent belief state full stop and how they could do so rationally. I discuss how a rational agent could end up in an inconsistent belief set in chapter 3, defending myself from a possible objection from Niko Kolodny (using a view of rationality as correct response to different pressures) along the way. I argue for the existence of inconsistent beliefs full stop in chapter 4; inconsistent beliefs are possible on an interpretationist framework despite some Davidsonian worries to the contrary.¹ In chapter 5 I pause to explore what these results mean for our understanding of human reasoning, including: the value of consistency, integrating different methods of reasoning, social reconstructions of reasoning, and how we should be evaluating agents.

Before moving on to the arguments of the dissertation, I will take another moment to clarify what I am and am not trying to do in this dissertation and make some terminological notes.

1.2 Notes on methodology

There are many different areas of study and sub-disciplines that deal with reasoning. Within philosophy alone, we find practical reasoning, epistemology, some areas of philosophy of mind and the study of formal systems like logics and decision theory. My project is an epistemological project that is trying to be responsible to philosophy of mind, but I do not shy away from dealing with practical reasoning as it comes up. (When I say I am being

¹While it may seem as though I should deal with possibility before rationality, particularities of Davidson's position make it easier to discuss after we discuss rationality.

responsible to philosophy of mind, what I mean is that I am trying to be sensitive to the ways that people really reason, so I mean it in a broad sense that includes much of philosophy of psychology.) I am focusing on theoretical reason as opposed to practical reason because several of the issues I want to discuss are clearer if we focus on the way that beliefs are related to one another rather than focusing on actions and how they fit into an agent's goals or life projects. That being said, many of the examples that I discuss, as you have started to see above, are closely tied to situations where an agent is also deciding on a practical course of action. This is partially because it is often when someone is thinking about what to do that they need to get clear on what they believe about complex issues like the ones I will be discussing.

Given what was said in the previous section about classical logic, I cannot apply it in a straightforward way to reasoning processes, e.g. by saying that it is irrational to be inconsistent according to classical logic or that it is rationally permissible to make any inference allowed by classical logic. (The latter would imply that it is rationally acceptable for an agent with inconsistent beliefs to infer any belief whatsoever.) A common response to the belief that classical logic is not a good guide to rational belief, is to look for another logic or formal system to serve as guide.² I want to make it clear at the outset that I am going to be setting aside formal models of reasoning altogether for my current project. Formal models of reasoning are one important tool among many, but they are not going to be useful here. This is because there are issues with global consistency constraints in general, and so advocating for an alternative way of setting those up will not be helpful in the current project. While there are other uses for formal models, I would like to make some progress on discussing how people do and should balance different pressures while they are reasoning

²For example, many-valued logics can be used when we want to model situations where some statements are both true or false, or possibly lack a truth value altogether. Fuzzy logics can be used to deal with Sorites paradoxes. And various sorts of paraconsistent logics can be used for those concerned with explosion. LP (Priest 1979) is one paraconsistent logic which is a variant of many-valued logics. Relevant logics are designed to avoid some paradoxes of implication, for example that an argument taking an arbitrary proposition as a premise and an unrelated tautology as a conclusion is valid in classical logic. These logics are also paraconsistent (see Mares 2004 for one discussion).

before attempting to formalize anything. Jumping too quickly into formalization can be detrimental unless the proper ground-clearing has taken place.

For the purposes of this dissertation I am simply trying to show that inconsistency is not necessarily irrational and point out some other features of reasoning processes. While obviously I cannot recommend classical logic as an acceptable way way of reasoning in all situations, I leave development of alternative formal systems (or other systems of reasoning) for another time and place.

Occasionally in the dissertation I will comment on issues surrounding credences or graded belief, but for the most part I will be focusing exclusively on all-or-nothing belief. This is primarily simply a result of what I personally find to be philosophically interesting, but I will make some brief comments here on why this is the case. First and foremost, I will be talking quite a bit about what how people would describe their own reasoning processes, and how we attribute beliefs to ourselves and others in everyday settings. We typically do not engage in this activity with the precision called for in a Bayesian epistemology. I think that this type of precise description has its time and place, but I am interested in seeing what is philosophically interesting in more typical ways we have of talking about beliefs, reasoning processes, and other mental states. Secondly, I want to address some common arguments I have run into against all-or-nothing beliefs being inconsistent. (And I do not have as much interest in discussing things like “Dutch book arguments” or representation theorems.) Some of the authors I discuss are also focused on all-or-nothing beliefs in particular, including Kolodny, who is the focus of chapter 3. Lastly, several of the questions simply seem cleaner or more interesting to me in the standard framework as opposed to the Bayesian. It is more of a mystery, for example, how people can arrive at contradictory beliefs than how people can arrive at credences in P and $\text{not}P$ that do not quite add up to one. How people should behave in order to reflect credences also seems to be a foggier issue than how they should act in light of their all-or-nothing beliefs.

I will end this chapter with a few terminological notes, starting with “reasoning.” Instead of attempting to give a definition, I will give a rough characterization of what reasoning is and discuss some benefits and drawbacks of this characterization. Reasoning is a primarily mental activity. It is typically described in terms of transitions between (sets of) propositional mental states, such as beliefs, ends, desires and intentions. These transitions are typically rule-governed. One benefit of this rough characterization is that it does not press me to take a stance on whether the reasoning process is conscious or unconscious. (I certainly do not want to say that reasoning must only consist of conscious phenomena.) Also, limiting the scope of reasoning to propositional attitudes allows us to rule out instances of lower-level brain processing. So, for example, processing in the visual system will not count as reasoning. But, in limiting it to just propositional attitudes, we may be ruling out some mental activities that we may want to call reasoning, especially if many mental states we use in reasoning are not propositional in type (like, say, mental models, mental maps, mental images, etc.). A discussion of whether reasoning with mental maps counts as reasoning with propositions or not would take us too far off-topic. I will be discussing mental maps later, so I would like the characterization to include mental maps but not include lower-level mental states. One may also want to push against the rule-governed part of this definition. I do not think that all reasoning can be described as goal-directed in any straightforward way. Also, less structured changes in mental states such as gestalt shifts might play an important part in some reasoning processes.

In this dissertation I will be using the term “rational” to refer to good reasoning. So, when I say that inconsistency can be rational, then I am saying that inconsistency can be a part of good reasoning processes or reasoning executed well. Here I will be discussing epistemic goodness or badness, though these issues are not unrelated to other ways of deeming reasoning good for bad (like practical benefits or drawbacks).³ While my thesis is that it can be rational to believe inconsistent propositions, I take myself to be engaged in an equally

³See Harman (2004).

descriptive and normative project overall. A normative theory that is in no way responsive to descriptive facts is in danger of being irrelevant to the system one is trying to theorize about. But normative facts about a system will be an essential part of a description of how a system operates. (For example, a broken calculator is incompletely described unless one mentions facts about what it should be doing and is failing to do.) And, in the case of reasoning in particular, if what we should be doing as reasoners is getting by in the world or tracking the way that the world is (or whatever one takes the primary epistemic norms to be), then a description of the system will not be complete without mentioning these facts. In this dissertation I will be focusing on the case of inconsistencies in a reasoning process. So, in my discussions of how inconsistencies play into reasoning, I hope to address both issues of how inconsistencies actually play a role in reasoning and how this makes sense given what reasoners should be doing.

Finally, I have a note about the term “inconsistent.” For the purposes of this dissertation, I will say that a set of beliefs is inconsistent if it does not have a model (in classical logic). This elides several complications regarding applying the idea of consistency to a natural language that will not matter for the arguments presented here. Definitions of when propositions, sentences or beliefs are consistent will typically be either syntactic or semantic.⁴ A set of propositions is semantically inconsistent if the propositions cannot all be true together. If we use this sort of definition in a natural language, it makes discussing dialethism more difficult, as dialethism holds that some sets of propositions that do not have a model in classical logic (or cannot all be true together in classical logic) can in fact all be true together. And a syntactic definition for a larger set of propositions is hard to pull off without a proof theory. So, I will stick with the definition in terms of classical logic, despite the fact that in some cases there could be issues surrounding how to express beliefs or statements in the language of classical logic. (For instance, we need to make sure that we do not hide inconsistencies

⁴Of course, many older definitions of the principle of non-contradiction, for example, are stated in more metaphysical terms. For example: “the same attribute cannot at the same time belong and not belong to the same subject in the same respect” (Aristotle, *Metaphysics* 1005b19–20).

with a misleading translation scheme. “My cat is black” & “My cat is not black” should be translated as “P” & “not P” not “P” & “Q”.) In this dissertation I will be focusing on cases where it is uncontentious whether the set of beliefs is inconsistent, so this issue will not affect my arguments.⁵ For the entirety of the dissertation, I will exclusively be talking about synchronic inconsistencies. While some consistency norms can be diachronic in nature, e.g., Bayesian updating, I am here concerned with synchronic consistency norms. I will sometimes say “contradiction” when I am referring to a set of two such propositions or say that a single belief is contradictory if it is of the form “P and not P.” “Coherence” is a less precise term used to mean that things fit together. I will use it when called for by standard usage in the literature (so, when discussing probabilism and coherentists, including Davidson), or when a more broad or less precise sense of “fitting together” than consistency is called for.

In the next chapter, I will start by discussing one common way of dealing with inconsistent beliefs, fragmentation.

⁵For a survey of definitions of consistency, see Grim (2004).

CHAPTER 2

Isolating inconsistencies

In this chapter I discuss the fragmentation view, which is one way of handling a conflict between common assumptions built into theories concerning rational human agents on the one hand and actual inconsistent agents on the other. The fragmentation view is intended, in part, to explain how an agent could plausibly have inconsistent beliefs, especially on a view of belief that includes classical logical closure. It is also intended to make sense of how in some cases we can consider an agent with inconsistent beliefs to be rational. I will start by reviewing the motivations for and commitments of the fragmentationist view, and I will then argue that this view cannot handle some cases of humans reasoning with inconsistent beliefs. The focus will be on how fragmented agents actually use beliefs in reasoning. With this in mind, I will discuss the manner in which I think it is helpful to divide belief systems into subsystems and how agents should be using their fragmented belief systems more broadly. I will finish up by locating some objections to the alternative view I have sketched as a way of setting up the rest of the dissertation. Seeing how we need to modify the fragmentation view will be useful for seeing how a particular type of human reasoning works: reasoning with conflicting ways of thinking about the world. I will conclude that in some situations, the rational thing to do is to reason with two belief subsystems at once, ones that may be inconsistent with one another. This does not mean that the idea that it is useful to think of human's belief systems as broken into subsystems should be abandoned.

2.1 Fragmentation

Human beings occasionally find themselves with inconsistent beliefs. This causes problems for many commonplace ideas about belief. In this dissertation I will be focusing on two in particular: First, the idea that we should, when attributing beliefs to someone, avoid attributing beliefs that are inconsistent. This idea can take the form of denying that people can really believe inconsistently or that it is irrational for them to do so and we should be more charitable than that. Second, the difficulty in seeing how an agent can effectively make decisions or form new beliefs if their existing belief set is inconsistent. The rest of this section is dedicated to explaining a very common response to these problems: fragmentation.

A simple way of thinking about belief attribution is as follows. People do what they think will get them what they want, based on what they believe. More generally, if an agent believes P , then they will act as though P is true (I will sometimes call this the classical view). Given this, it is hard to make sense of how we could attribute inconsistent beliefs to someone. How can an agent in a single action be acting on two beliefs that directly contradict one another? It is easier to make sense of an agent sometimes acting in accordance with one belief and sometimes in accordance with another belief that contradicts it (Egan 2008, 50).¹ This issue surrounding belief attribution is one motivation for the fragmentation view. As a first pass, the fragmentation view states that agents do not have one consistent belief system, but rather different internally consistent subsystems that they use at different times. Each individual subsystem must be internally consistent, but they do not need to be consistent with one another. I will, to fall in line with the literature here, call these consistent subsystems “fragments.” To clarify the view, Seth Yalcin (2016) has used the metaphor of having an atlas with which to navigate the world as opposed to a single map. To be useful any single map must be consistent, but the maps need not be consistent with

¹Davidson, of course, has a much stronger version of this sort of belief attribution worry. Given his version of principle of charity and his way of approaching radical interpretation, inconsistent beliefs pose a serious challenge for him that he meets by saying that an agent can never be aware of two contradictory beliefs that they hold (see 2004a&b). I discuss Davidson in more detail in chapter 4.

one another. Whenever an agent needs to think through a theoretical or practical problem, they will use one of the maps they have at their disposal which is well-suited for the purpose at hand. We should expect that a lot of information will appear on multiple maps, so, while we do not have a solid grasp yet on what is involved in saying that the maps are different, it is important to note that it does not necessitate that the belief sets are disjoint.² Davison notes this explicitly,³ but it is also obvious that no matter what your more specific beliefs are, you will need basic beliefs about things like the structure of space and basic physical laws more or less all of the time.

Another motivating worry for fragmentationists is that it is hard to see how an agent can reason using an inconsistent belief set. This worry is motivated at least partially by the idea that human reasoning has some relationship to logic. If agents are permitted to infer the logical consequences of their beliefs, then anyone with an inconsistent belief set is permitted to believe anything at all.⁴ Even without the commitment to this way of thinking about the relationship between classical logic and belief revision, many still think that it is hard to see what basic inferences one will or could make on the basis of P and not-P. If the information of whether P is relevant (and why else would it be brought into the reasoning at hand), then believing that P and also believing that not P is not helpful at all. Having just those two beliefs is not any better than starting from scratch.⁵ Fragmentation can deal with this sort of worry nicely. If one is fragmented, one can only infer on the basis of a consistent subset of their total belief set. So, one will never be confronted with the problem of figuring out

²I mean “disjoint” in the set-theoretic sense of having no elements in common.

³Of course this is important for Davidson because of his holistic views. Again, I will have more to say about Davidson in chapter 4.

⁴This is because of explosion. B is a logical consequence of (A & - A) for any propositions A and B. (And if a set of propositions is inconsistent, then you can derive some sentence of the form (A & - A)).

⁵Of course this is focusing on one particular kind of inconsistency. For the majority of this chapter I focus on directly contradictory beliefs instead of larger inconsistent sets because they seem to be the harder case. People tend to have less of an issue imagining how someone could function if they had beliefs that were inconsistent in the style of a preface paradox, for example.

what to believe in light of a contradiction. Fragmentationists, who want to preserve classical inferences as much as possible and avoid some types of paraconsistent logic (see esp. Lewis 1982), find this property of the view attractive.

The last motivation for fragmentation I will discuss is an issue of memory access. There are many different reasons that one might want to divide beliefs up depending on how one tries to access them. One is related to this issue of consistency; many people find it more plausible that agents can have inconsistent beliefs if they are unaware of the inconsistency. So, if an agent can only access one consistent subset of their beliefs at a time, this makes the idea that people have inconsistent beliefs more plausible.⁶ But there are also considerations unrelated to the possibility of inconsistent belief.⁷ Elga and Rayo (2021) discuss many simple cases to motivate the need for an access indexing on beliefs. An illustrative one is the case of Jack, who remembers when asked “who is Beatrice Ogden?” she is the person who lives in apartment 2h, but when asked “who lives in apartment 2h?” would be unable to tell you. In this situation the question “does Jack believe Beatrice Ogden lives in 2h?” is overly simplistic and not a helpful way of getting at what his belief state is. The information about Jack’s belief states is more complex than that type of question would allow. Elga and Rayo use these considerations to argue that we should think of beliefs as being relative to elicitation conditions. In the example above, Jack has the belief that Beatrice Ogden lives in 2h in one elicitation condition but not the other.

It is worth mentioning here that not all fragmentationists are particularly concerned with issues of belief. Another related branch of fragmentationist thought is simply concerned with working out the formal issues of how to deal with an inconstant corpus that keeps as close to classical logic as possible. Schotch and Jennings (1980) examine the problem of what to

⁶This is, roughly, Christopher Cherniak’s motivation. See his 1981, 1983, 1986. He is concerned about being able to attribute beliefs to agents without assuming that they are ideally rational or failing to require them to make any inferences whatsoever on their existing beliefs.

⁷Yalcin (2016), following Robert Stalnaker (1984), discusses the nature of belief as being intrinsically question-sensitive. This view is motivated by many things, one of which is concerns about logical closure of belief. (I will not be discussing this issue here.)

do with an inconsistent data set, since that is sometimes what is available.⁸ They take it as obvious that people do have inconsistent beliefs that we are aware of (and do not normatively comment on this fact) and instead busy themselves with explaining what we should do once we find ourselves in such a position. But their answer to the question of what to do with an inconsistent data set is that only consistent subsets of the set of available propositions should be used as premises in a single proof so that classical rules of inference can be used.

Given these motivational concerns, what will a fragment look like in a human agent? For the most part, these authors do not say much about what a fragment amounts to beyond the discussion of a few examples and the idea that a fragment must be consistent. Two exceptions to this are Davidson and Elga & Rayo. For Davidson, the fragments are caused by an agent having an irrational reaction to an uncomfortable situation. For example, the pain that someone feels due to the realization that they are going to fail a test might cause them to think that they will not fail it (2004a). Because this new belief contradicts an already existing belief, it will have to inhabit a new fragment. The boundaries between the fragments are not permanent, the agent can dissolve the boundary by becoming aware of contradictory beliefs on either side of it. He also makes a point of saying that the boundaries are just conceptual aids for us to understand the agent and are not things that we should expect the agent to be able to detect via introspection (2004a, 211).

The Elga and Rayo way of making sense of a fragment relies on the idea of access; two mental states are in the same fragment if they can be accessed via the same elicitation condition. They give a filled-out version of a classic case due to David Lewis (1982) in order to elaborate on this idea, where Andy thinks contradictory things about the orientation of a road depending on how he is moving about the city. When Andy is on foot, he thinks that Charing Cross road runs East-West because of how his mental foot-map is organized. When he is on the train, he thinks that it runs North-South because of a misleading turn in the train's route. This influences him to behave in some ways that one would expect. If trying to

⁸See also Rescher and Brandom (1979).

walk to Charing Cross road, he will take a street that runs North-South. If someone asks him about Charing Cross road on the train, he will give answers to consistent with it running North-South. He is not one to reflect much on the layout of his surroundings, so he has never noticed this inconsistency. This example described in this way makes minimal changes to classical ideas about how beliefs and actions are related, it simply adds a situational indexing. There is no point at which Andy is going to be acting as though he thinks that ⟨Charing Cross road runs North-South and East-West⟩. Nor will he be acting on two beliefs at the same time: ⟨Charing Cross road runs North-South⟩ and ⟨Charing Cross road runs East-West⟩. He sometimes acts as though he thinks it runs North-South and sometimes acts as though he thinks it runs East-West. When Lewis discusses the original version of this example, he says that the person will never reason with an inconsistent set, and if they become aware of the inconsistency in their beliefs, they will fix it. Elga and Rayo use the example to show a realistic case where someone might have different beliefs that are activated in different situations. It is realistic to think that one might have a slightly different mental map depending on how one is traveling.

So far, I have canvassed some distinct views under the umbrella of “the fragmentation view.” In this chapter I will be focusing on two specific aspects of this cluster: one I find to be a helpful expansion of the classical way that we think about beliefs and one that I think is mistaken. First, I agree with the idea that we should not view agents as having one large consistent belief system, but rather as having several belief subsystems that are not necessarily all consistent with one another. This idea is common in one form or another to all fragmentationist views. I will suggest a way to start filling it out. Second, I will be arguing against the thesis that an agent cannot use more than one fragment in a given episode of reasoning. I will show that agents do in some circumstances use more than one fragment in a single episode of reasoning. Not only that, but they are rational to do so. This part of my argument is only targeting the fragmentationists who are motivated by consistency worries and not by access worries. (If an agent uses different beliefs in a single episode of reasoning,

they will by definition be in the same access-fragment, so there is nothing to argue about here.)

There is one important element of these theories that I have not yet directly discussed. I have phrased the main claim that I am arguing against as “agents do not use more than one fragment at once.” But, in the case of some fragmentationists it may be more accurate to say that they are claiming that a *rational* agent will not do so. Davidson thinks that you cannot, even irrationally, hold on to inconsistent beliefs once you are aware of them; an agent can never be aware of the two at once (2004a, 200). This means that we will not be able to reason using different fragments together. Other fragmentationists will be less radical in this respect and will think that an agent can be aware of different fragments but will not be able to actively reason on the basis of more than one at once. (Schotch and Jennings say this when they speculate about the possible application of their theory to the case of a human agent.) Lewis says that an agent *should* not reason using beliefs from different fragments (1982, 436) and, in addition, since he is talking about logic, it seems as though he is talking about idealized agents. Stalnaker similarly seems to be talking about idealized agents (1984, 1991). This distinction among fragmentationist views will not make a difference to my argument here because I will be arguing both that agents do use multiple fragments and also that they are right to do so.

So, for the rest of this dissertation, I will take the fragmentation view to be the view that it is helpful when theorizing about agents to divide up their beliefs into fragments and agents cannot use different belief fragments in a single episode of reasoning.⁹ Fragments must be

⁹Daniel Greco (2014 and 2015) advocates for a different type of fragmentation that is not a division in the mind among beliefs of the same type, but a division separating one type of belief from another. (Kindermann and Onofri [2021, 13-14] call Greco’s style of fragmentation “vertical” fragmentation and the type of fragmentation I am discussing in this chapter “horizontal” fragmentation.) It is motivated by purported counterexamples to the principle that if one believes a proposition, then one believes that one believes it. While it would be an alternate way of handling inconsistencies, it is not equivalent to fragmentation as defined here and is not directly motivated by inconsistencies. I will briefly review why.

One can certainly hold that there are two types of belief-like states without thinking that they are not used in the same reasoning process or that one cannot even be aware of them at the same time. Take implicit-type belief as one example: You can certainly be aware that you have different beliefs and beliefs

internally consistent but need not be consistent with one another. Different fragments can have overlapping beliefs, so an agent is allowed to use beliefs that are in fragment B while reasoning with fragment A, as long as it is a belief that is also in fragment A.

There are a couple of reasons why I have decided to characterize the fragmentationist position in terms of episodes of reasoning. First, in this chapter I am concerned primarily with issues that an agent may run into while trying to function with inconsistent beliefs. Second, other characterizations in terms of different contexts or different situations do not seem to be as accurate.¹⁰ Let us take a slightly changed example from Andy Egan to show this point (2021, 122). Roy is an excellent chess player but is bad at explaining chess strategy. So, he can be playing an excellent game of chess, but be unable to explain why he is doing what he is doing as he plays. This is certainly a singular situation. It is arguably a different context, but I believe it is more theoretically clean to divide by episodes of reasoning, since clearly in this case the chess playing and the explaining of the chess strategy are actions that are coming out of different reasoning processes. One more clarification is needed on this characterization before we move on to my arguments against fragmentation. It is important to note that Davidson (2004c, 182) says that the different parts of the mind must have been in contact with one another to generate the belief that one action is better than another all things considered. But he does not say here that the different parts of the mind have a direct contradiction between them. So, Davidson is someone who believes something similar to me in thinking that there are parts of the mind that are semi-autonomous, but we can consider them both at the same time. What he thinks cannot happen is that you consider

(to use terminology introduced by Gendler 2008). Even if you think that the “access” will be relevantly different, they can still both be used in the same reasoning process in some sense. Classic examples of someone reasoning about something touched on by an implicit bias of theirs will all be of this type.

It is more obvious that it is possible to be a fragmentationist as defined in this chapter without being a Greco-style fragmentationist. One can think that beliefs of the same type cannot be consciously accessed at the same time. Also, it can be that you have beliefs of the same type that cannot all be in the same fragment. So, while it may be that having the beliefs in different fragments be of different types would help us to fill out a story of what a fragment is, that is an independent position than the one detailed being discussed here.

¹⁰Kindermann and Onofri characterize fragments in this way (Kindermann and Onofri 4-5).

two things at once where there is an actual *contradiction* when you take them as a whole. The fragmentationist position is a moderate one between the extremes of denying that there is any way that a rational agent can have inconsistent beliefs and denying that there is any kind of problem with inconsistent beliefs. My position will be at most half-fragmentationist, since I will agree that talking about belief fragmentation is helpful but I will disagree about the stringency of the boundaries between the fragments.

2.2 Against fragmentation

In this section I will lay out my objection to the fragmentation view. I will start by setting my objection apart from an objection due to Aaron Norby (2.2.1). I will then discuss some important differences between my approach to how to think of a belief fragment and the fragmentationist's approach (2.2.2). In particular, I want to look at different sources of inconsistency and fill out what a fragment or other belief subsystem might look like. Next I will discuss a detailed case, the case of making a resolution "against the evidence" (2.2.3). In these cases, it seems clear to me that the best option is for the agent to have inconsistent beliefs, but also these beliefs cannot be part of separate non-interacting fragments in the mind. I will then step back from this particular case to discuss what this means for the fragmentation theory and also for the structure of reasoning generally (2.2.4).

2.2.1 Norby against fragmentation

Norby has argued that the fragmentationist theory is explanatorily bankrupt (2014). He starts by summing up the classic view relating beliefs to actions, that our beliefs are straightforwardly reflected in our actions. If an agent believes P, then they will act as though P is true. But, of course, the fragmentationist cannot be committed to anything that simple. Andy believes that ⟨Charing Cross road runs North-South⟩ all the time, but he does not act as though he believes it when on foot. Whatever theory they have that relates beliefs to

actions needs to make reference to a subset of beliefs, since they think that it is possible for an agent to have inconsistent beliefs, but impossible for inconsistent beliefs to be together reflected in an action. A problem arises here because in order to use this way of relating beliefs and actions for prediction's sake, the fragmentationist will need to be able to say which fragments are active at a given time. They may look to empirical psychology for help here, but the fragmentationist cannot employ any of the standard ways that psychologists talk about these phenomena. This is, roughly, because they all involve the beliefs active at different times having different subject matter and the fragmentationist is committed to beliefs with extremely similar contents being in different fragments. Most of their cases will involve beliefs that are about exactly the same things, as exemplified by the mental map case above. This means that they have no way of saying which beliefs are active when. So, Norby argues, the fragmentationist is committed to saying something very trivial, which is just that the active beliefs influence actions, but the inactive beliefs do not, with no indication of why some are active, and some are not. This means that they are not doing much beyond re-stating the original problem that they wanted to explain, which is that people can have inconsistent beliefs.

I agree that the fragmentationist is not helping to explain how inconsistent beliefs can exist in an agent and rationally cause behavior. I am not sure, however, if this will be a fatal blow. A fragmentationist could respond to this objection by trying to fill in their explanatory picture. They were initially motivated by basic threats that come from the existence of inconsistent beliefs and could expand the picture to include ways to predict behavior. Cherniak, an early fragmentationist, suggests looking to the study of memory to fill out the picture (1981, 1983, 1986). I will go further than Norby and claim that even if they did have some way of filling out their story of when a belief fragment is active, they would still be unable to give a satisfying explanation of how beliefs and behavior are related in a fragmented agent. This is because the issue with the view is deeper than just an underdeveloped sense of what counts as a fragment and what determines when a fragment

is being actively used by an agent in reasoning. In fact, the idea that an agent will only be able to use one fragment in a given episode of reasoning is flawed from the start. I will now turn to a discussion of what reasoning will look like in a fragmented agent before getting to the heart of my objection.

2.2.2 Fragments and belief subsystems

Human reasoning is not a monolithic phenomenon. How human beings come to decisions about what to do or what to believe differs depending on situation, subject matter, what kind of information is available, etc. This fact should not surprise us. We have a lot of different ways of putting information together and coming to decisions, many of which have been patched together as our environmental niche has changed and become more complex. And, indeed, we have psychological evidence that humans have many different ways of reasoning, including a menu of different types of heuristics.¹¹ In short, we have many different reasoning methods available to us.

In this chapter I am interested in exploring the useful and misleading parts of the fragmentation view. I believe that the idea of fragmenting mental states is a useful one, but I disagree about the level of separation needed among the fragments. In this section I will fill out the idea of a fragment, in order to give some more detail to what sort of mental separation I think is useful. As a first pass, let us say that a mode of reasoning is a collection of rules and habits that have been built up around a particular kind of common reasoning problem. It can be helpful to talk about belief subsystems associated with a mode of reasoning: those beliefs that are used by or generated in a particular mode.¹² A belief may be in multiple such belief systems. There is some pressure for such subsystems to be consistent, as it is easier to deal with a set of propositions if they are consistent and the beliefs in the

¹¹See Gigerenzer et al. 1999.

¹²Nicholas Rescher utilizes a somewhat similar distinction, but for sources of information as opposed to methods or modes of reasoning (1988, 5.3).

subsystem will be subject to a lot of the same pressures. (And if they are, we can call them fragments.) All of this will become clearer when we get to discussion of some examples.

Navigation as a whole is one example. When navigating an area, we have a mental map that we use and update. We gather information by asking around, looking around and consulting nonmental maps. We integrate information from these different sources and place ourselves on our mental map via landmarks, major streets, etc. These mental maps do not need to be extremely detailed, but they should be consistent and have enough landmarks in order to be usable.

These belief systems that are associated with reasoning methods will sometimes be fragments (when they are consistent), but not all fragments summoned to explain cases by fragmentationist will be neatly explained by a subsystem associated with a mode of reasoning. Take the Andy example. One can choose to say that there is a different mode of reasoning associated with train travel than with navigating on foot. But there do not seem to me to be enough differences in what you are trying to do or what the rules for success are to count as being a different mode of reasoning.¹³ This is fine for my purposes, because I think this tracks an important difference between types of belief subsystems. Andy here is making a mistake in having inconsistent beliefs about the layout of the area. It may not be an important mistake, or even one that I would recommend rectifying, but it is still certainly a mistake. Whereas the examples I will get to below are ones where I think that the inconsistencies coming from different subsystems need not be fixed, and this is because they are from different modes of reasoning.¹⁴

Another example (that will be useful shortly) is taking on testimony. People tend to take people at their word when they report facts, but we also have a lot of default rules for when we should be more critical and when we should treat someone as an expert. We will tend

¹³And, on the other extreme, we do not want to say that something too broad counts either, like, “deciding what to believe,” as an extreme example.

¹⁴Or more different, if you are comfortable saying that train travel and travel by foot are different modes.

to take others as experts on their own beliefs. More generally, If one person I am talking to has a lot of training in an area and the other does not, I will tend to trust the person with training. We can also have beliefs and rules that are specific to certain people (Joey tends to lie, Clarissa is very trustworthy, Evan is a reliable source of information about cars but not about motorcycles, etc.).

So, this idea of belief subsystems associated with modes of reasoning is another way that we can talk about mental divides more concretely. This way of talking is different than the standard fragmentationist versions. We do not have a detailed picture, but we are starting to have more to say about them than that they are active at different times. They are not caused by any avoidance of the truth, but rather by different ways of reasoning, following different rules or norms, etc. This particular way of thinking of mental divides is helpful for saying why in some cases it can make sense to have inconsistent beliefs: the different beliefs are the result of following different rules and were formed in importantly different ways.

Given that we have all of these ways of reasoning about the world, we should expect that for complicated questions we will often be confronted with different answers, even contradictory answers, to a single question from different reasoning modes. Unlike the case for beliefs that were generated by the same methodology or following the same rules, we have no reason to think that these beliefs will be consistent with one another. Different modes of reasoning will often have overlapping subject matters (or be neutral with respect to subject matter) and we would not expect that different methods will always give the same answer to a question.^{15,16}

¹⁵We will see Kolodny give an argument that is at odds with this idea in the next chapter, which I will address.

¹⁶Rescher makes a similar point but regarding sources of information (1988, 79). He suggests that in many cases when confronted with inconsistent information from different sources it is best to preliminarily accept it and go from there. I am suggesting that in some cases, specifically when the inconsistencies occur across modes of reasoning with entirely different standards, one should more than provisionally accept them.

A fragmentationist would have no reason to disagree with anything said in this subsection so far. They will object, however, to my claims regarding what happens when an agent does have inconsistent beliefs in different fragments associated with different modes of reasoning and is confronted with a problem. The fragmentationist will have to say that one of two things must happen. The agent must either only use one of the fragments, or they will have to defragment. Defragmenting involves erasing the boundary between the two fragments and resolving the inconsistency, which will mean erasing at least one belief. And because there are connections among beliefs in a mode of reasoning, changing one belief may necessitate changing a lot of others.

I will be arguing that using only one fragment leads to an impoverished method of reasoning. The defragmentation option will also lead to issues in the cases that I am interested in considering, because in the process of defragmenting the original assumptions and methodologies of the different reasoning methods may also have to be altered. This is because a developed mode of reasoning there are a lot of interconnections among the beliefs. We want the mode itself to be consistent and have connecting threads of various kinds. You cannot just axe a few beliefs out here and there. This is not a desirable result if each of the methods have independent value that would be lost in the process of defragmenting. I am interested in exploring an alternative picture on which the fragmentation within an agent is less severe.

The rest of this chapter will proceed as follows: I will walk through one case in detail (and suggest some other cases where a similar treatment may be illuminating) and then I will sum up what I believe this case says about the fragmentationist position. Lastly I will address some objections and replies and set up my tasks in the rest of the dissertation.

2.2.3 Resolving

Let us start with the basics of a case where one is reasoning about oneself in a difficult situation. One has made a resolution to not smoke, but one has made such a resolution several times before and failed each time. (The choice of quitting smoking is arbitrary. One

could substitute in any number of different types of resolutions that people make. The only important property of the example is that it is one in which people will feel a strong temptation to break the resolution.) In a case like this an agent will often need to have beliefs about whether they themselves will or will not keep their resolution. Maybe they need to decide whether to give away their cigarettes. Maybe they need to decide whether they will promise to someone else that they will not smoke. But this presents a thorny epistemic situation. Roughly speaking, one can believe that one will smoke again (and not that one will not), one can believe that one will not smoke (and not that one will), one can have both beliefs, or one can have neither. All of these states seem to exhibit some normative defect. I will now go through the options in more detail. This discussion is, of course, assuming that one can resolve to do something that one has failed to do in the past. One could avoid this epistemic situation by never resolving to do something in the face of having failed at it many times in the past. But I take it that we do not think that the only rational response to failure is to give up.

First, you can believe that you will smoke again.¹⁷ This seems rational insofar as that is the conclusion that a disinterested induction would support. Meaning, it is the result you would get if you did a simple induction on similar situations in your past. Of course, if this is your belief and you act on the basis of it you will run into trouble. If you are going to smoke you should buy cigarettes, you should go out for smoke breaks with your friends, etc. This means that it was pointless to make the resolution in the first place. Of course, alternatively you could have the belief but not act on it. But this is a basic form of irrationality. One of the two core functions of beliefs, along with representing the world, is to be a guide to action. If a belief is not in any way reflected in the actions of an agent (and no strange circumstances are present in the vein of super-Spartans)¹⁸ then it is hard to say that the agent has the belief at all. This way of reacting to the situation is also bad in that it does

¹⁷Richard Moran argues that this sort of thinking is always a mistake, see his 1997 and 2001.

¹⁸See Putnam (1963).

not take into account an important part of the situation: that agents have a particular kind of control over their actions.

Second, one can believe that one will successfully quit smoking, “against the evidence.”¹⁹ This is what Marušić claims one should do (2015). This makes sense from the agential mode of thinking. If I decide to do an action, I will do it. I am an autonomous, efficacious agent. Without having this mode of reasoning at one’s disposal and implementing it, it is hard to see how one could function as an agent at all. This point of view is necessary for planning and acting. I cannot really plan to do something unless I think I can do it. In this example, the agent plans on not smoking. In order to plan to do something they must think that they will do it. But it seems like in this case the agent is not taking into account all of their observational evidence.

Third, we have believing neither proposition. But remaining agnostic on a question like this does not look like a good option. If one does not have any beliefs whatsoever about the questions, it is not clear how one can make decisions like the above. I am going to table a more thorough discussion of this option for now. How high the threshold for belief is, the circumstances under which agents should suspend belief, and why we need beliefs in these types of cases specifically, will be essential for my argument in the next chapter and will be discussed there. In this chapter I am focused on arguing that the fragmentationist story of how an agent can function in an inconsistent belief set will not work.

Of course, there is another way that one could believe neither thing while not remaining agnostic. That is if one did not have full belief in either the proposition that one will smoke or the proposition that one will not smoke, but rather had credences in the two propositions that sum to one but are both low enough that we would not say that one had a full belief in either. I do not have space to fully deal with the issue of credences in this chapter, but

¹⁹I put “against the evidence” in quotes because I do not want to take a position here about the accuracy of this way of thinking about these cases, but this is how they are referred to by Marušić and others in the ethics literature. Marušić himself notes that “despite the evidence” may be more descriptively accurate, because it is not necessarily the case that the evidence decisively supports believing that one will not succeed.

I will quickly mention why I do not find them to be a promising way of dealing with this problem. First, it is not clear how having such credences would be the appropriate response to the evidence at hand. Second, it does not seem as though having such credences will help the agent know what to do in this situation.

This leaves believing both. That option sounds bad because it means that the agent will be believing two propositions that contradict one another. I am going to be spending most of the rest of the dissertation arguing that this is not as bad as it sounds.²⁰ While there is certainly some pressure to avoid believing inconsistently, the nature of this pressure needs to be explicated. Then we can see that in some cases, like the case of resolving against the evidence, having inconsistent beliefs is the best thing to do.

I would like to start by noting that in both the case of only believing that you will not smoke and the case of only believing that one will smoke, the bad result came primarily from the lack of the other belief, not from the existence of the current belief by itself. The agent is not taking into account something important about the epistemic, practical or moral situation. I propose that the best solution to the problem is to form both beliefs. Of course, the fragmentationist need not disagree with me so far. They could simply say that the different beliefs are formed in different fragments and the agent does not use both of the contradictory beliefs in any particular decision or bit of theoretical reasoning. And the fragmentationist solution does seem to avoid some of the problems of the above options. The agent cannot be blamed for ignoring their observational evidence or for not treating themselves as an agent on the whole. The issue is that for any particular decision the agent will still be falling into one of the traps outlined above. To sum up, I agree with the fragmentationist that we can form both beliefs, but I disagree that the agent will be unable to function well if they can access both of them in a single episode of reasoning.

²⁰Pryor argues something similar about preface paradox-type cases in his (2018), but preface paradox cases are very different from the cases I am discussing here. Also, he does not discuss actively using an inconsistent belief set, he is arguing that ending up in an inconsistent state is the least bad outcome of some tricky epistemic situations.

To see why, remember that the fragmentationist cannot allow the agent to make a decision or perform a bit of theoretical or practical reasoning on the basis of both beliefs (the ⟨I will smoke⟩ and the ⟨I will not smoke⟩ beliefs). This means that for each individual decision the agent must either rely on one or rely on the other or defragment. So, if we look at the beliefs of the agent as a whole, it appears that they do not commit the mistakes of having only one of the beliefs, but these normative defects pop right back up again when it comes to evaluating particular decisions the agent is making or how they are forming particular new beliefs. Simply forming beliefs that respect one's agential status will not do any good if one is making decisions (or forming beliefs) that ignore these beliefs wholesale. Similarly, having beliefs that are responsive to one's observational evidence is not helpful if those beliefs are not taken into account in further deliberation and decision-making where it is called for.

Instead, what the agent should do is intelligently weigh the considerations on both sides. This will mean sometimes favoring one over the other. It will not mean deciding to throw out one perspective on the world.²¹ For example, should the person make a costly bet with a friend that they will avoid smoking for an entire year? Either decision one could make seems like it could be rational, depending on the circumstances. These can include how important it is that one stops smoking, how much of a blow it would be to lose the bet, and other similar factors. What does not seem rational is not taking both sets of beliefs into account when making the decision.

Let us take a closer look at Joseph, who is having a hard time quitting and ends up deciding to make a very expensive bet with a friend that he will not smoke a cigarette for the next month. It would be irrational of him to not take into account both the dispassionate evidence he has regarding how difficult it will be and the fact that he has control over his own actions. So he should not be making decisions while using a fragment of his beliefs that includes only one of ⟨I will smoke⟩ or ⟨I will not smoke⟩. Only reasoning with one, while disregarding the other exhibits the normative defects outlined above. The fragmentationist

²¹Thomas Nagel discusses a similar struggle in *The view from nowhere* (1986).

could try to carve out a fragment that includes some of the considerations from the dispassionate evidence mode of reasoning and some from the agential one. This feels ad-hoc to me. In the next chapter I will lay out my reasons for thinking that suspending those beliefs in the first place is not a rational response. And if they are beliefs held by the agent, a mental barrier between them and the considerations in favor of them does not seem to be in the spirit of the original fragmentationist project. I think it is helpful to think of divisions of belief in terms of modes of reasoning or even in terms of access, but arbitrarily cutting off beliefs from the reasons to believe them in the first place is another matter entirely.

Of course, there is a cost to this, which the fragmentationist is not willing to pay. On my view within a mode an agent should be consistent and can use classical inferences to reason. But when it comes to reasoning with more than one mode, agents do not have blanket rational permission to draw classical inferences of what they believe. If they did, they would be rationally permitted to believe anything at all. This raises the question of what the standards are for reasoning with more than one mode and what agents are allowed to infer. I am skeptical that this question can be given a thoroughly general answer. Questions about evaluating the rationality of agents and the costs of inconsistency will be discussed more in the next chapter.

So, we have one case on the table, but it is an open question how prevalent this complex type of balancing is in reasoning. In this subsection the central point I am making is that it can happen, so the fragmentationist is wrong that humans cannot function well with accessible inconsistent belief sets full stop. I am happy to concede that many examples unfold in the way suggested by the fragmentationist. But I am discussing this issue because I think that it points to something interesting about the structure of reasoning, so I will say a few things about how common it will be to use inconsistent beliefs from two (or more) fragments in a reasoning process without defragmenting.

To start, even in the case where you decide to act just on the basis of one of the fragments, the decision involves both fragments. Let us say that the sister of someone who has resolved

to quit smoking decides to act in accordance with her beliefs that the agent will succeed and set aside doubts based on past evidence. Here there are inconsistent beliefs at play in the reasoning process: she believes that her sister will quit smoking and believes that she will not. This is a very different case than if she had simply not considered the countervailing evidence, which seems like what she would have to be doing on the fragmentationist account. Remember that the core claim that the fragmentationist is making that I am objecting to is that an agent cannot use more than one fragment in a bit of reasoning. In order to consider the countervailing evidence, the sister would need to have considered beliefs from two different fragments. According to the fragmentationist, she believes both contradictory propositions, but one is in some way hidden from the reasoning process. The opposite type of case also occurs, deciding to act in accordance with the belief that one's loved ones will not succeed in what they are claiming they will do. People hit a point with others where they stop taking their proclamations as agents seriously. This is also a decision, one that would not make sense, and in particular would not make sense as a painful moment, if they just did not consider that they should generally treat others as agents.

I believe that there are other cases that should be handled in the same way. The conflict between the agential mode and the observational evidence mode is one that comes up a lot in decisions about how one should treat oneself and others. In addition to this, there are cases of everyday ways of thinking about things conflicting with more scientific or skeptical ways of thinking. What these cases have in common is a sort of incomparability of reasons. I will not be talking about these other cases in detail here, but this idea of incomparability will come up again in my discussions on the nature of inconsistency in the next chapter.

2.3 Conclusion

So, we can see that sometimes it is better to steer by more than one map at once. Having the benefit of two rich maps can be beneficial, and this benefit is not undermined by their not

being consistent with one another. While I agree that having nothing but a bare contradiction to work with *can* mean that the contradiction needs to be dealt with before moving forward, it *need* not if the branches of the contradiction come with more richly textured ways of looking at the world. If I am doing my taxes and come across some inconsistent information in my records I cannot move forward until I have resolved the issue. This is both because I need, practically speaking, to have the correct numbers before I move forward and also because the inconsistency is obviously because of a simple mistake made somewhere. If I am trying to navigate quitting smoking and I come across inconsistent beliefs the issue is more complicated. I can re-examine the evidence and other beliefs connected to each of the contradictory beliefs and consider which I should prioritize depending on the situation.

Remember that the fragmentationists are motivated by worries about how reasoning on an inconsistent belief set is possible and how it can make sense to ascribe inconsistent beliefs to an agent. (I am not presenting a formal theory here and nothing I have said here is inconsistent with the idea that there are some cases of information being variably accessible, so I do not need to worry about how to respond to those motivations.) In this chapter I have argued that reasoning with an inconsistent belief set is possible without fragmenting. In the next two chapters I will turn to how I will deal with the second worry. Before I get there I have one related objection that the fragmentationist could raise in response to the picture I have painted here.

The fragmentationist could push further and give another way of describing these cases that does not involve retaining both of the contradictory beliefs. Most likely they will try to give a description where those particular beliefs are withheld. This is how Egan would push us to describe these cases, given his discomfort with saying that a single action could be the result of reasoning on an inconsistent belief set. But I have a way of narrating these cases so that they involve inconsistent beliefs and have argued based on this narration that the fragmentation view causes problems. I will address this objection more explicitly from the point of view of choosing among different possible descriptions in chapter 4, where I

will talk about belief-ascription (which is the worry motivating the fragmentationist that we have not discussed in a great amount of detail in this chapter). In short, there is a way of describing these cases in which the set of beliefs that the agent is acting on is fully consistent, but that story is going to be missing important explanatory pieces. This is also the chapter in which I will explicitly address the fragmentationist's worries about ascribing inconsistent beliefs to an agent. I will be examining the original motivation for the consistency norm on belief attribution and showing how the splitting up of belief fragments I have postulated is sufficient to deal with the legitimate part of those worries without the full fragmentationist view that the agent cannot appeal to both in reasoning.

While not my main aim in this discussion, from here we can also start to see one reason that it is generally a good idea to avoid having inconsistent beliefs. Maintaining the inconsistencies comes at a cognitive cost in some cases. If I need to decide whether to throw out my cigarettes the answer is easy if I have one or the other of the beliefs in the set {I will never smoke again, I will smoke again} but it is more complicated if one must deal with both. In this case, one has to figure out how to integrate results from different modes of reasoning. This involves a lot of complicated weighing of different sorts of reasons. In fact, as I get to in the next chapter, part of what is so difficult in these cases is that there are incomparable considerations for an agent to deal with in their reasoning process, which greatly complicates things. Integrating the results will also involve figuring out how to respect the inconsistent propositions in one's theoretical reasoning and one's actions. These theoretical difficulties can trickle down into the practical and can be paralyzing. But, in some cases this extra work is worth it, because the complication of the situation demands it. This gives us a way of saying that certain cases of believing inconsistently are just the result of mistakes on the part of the reasoner or an issue arising because of bounded cognitive capacities. If the inconsistency does persist simply because the person has not considered the inconsistent beliefs all at the same time or has not executed a particular reasoning method well, then

we do not want to say that holding onto the inconsistency is rational.²² In order to argue for the rationality of the inconsistency above, we needed to tell a separate story for each of them that explains why it is rational to hold them. That means that it will not make sense to hold onto an inconsistency if all parts of it were arrived at in the same way and so are in the same fragment as I am conceiving of it. That would imply there was some sort of mistake.

I would like to take a moment here to sum up my results and to say something about what I think the fragmentation view gets right. Speaking metaphorically, what fragmentation gets right is that you need some distance among the inconsistent beliefs. In the fragmentation view this distance will be in the form of not being able to appear in the same episode of reasoning (and, according to some, they will not be able to fall under the awareness of the agent at the same time). What it gets wrong is that they need to be so far apart that they cannot both be used in the same reasoning event. Speaking less metaphorically, when one is reasoning in a particular mode, one needs to follow the rules of that mode. If I am in my lab gathering data, I should not record inconsistent things. I should make a consistent record of what I observe to the best of my ability. If I am trying to replicate the results of a previous experiment and I get wildly different results, I need to find an explanation for what is going on. If I am performing an induction on someone's behavior, I should either come up with a clear pattern or suspend judgment. But if I am zooming out and considering some lab results along with some skepticism about my methods there is no problem with continuing to hold onto beliefs that may not be consistent with each other.²³ The difference here is in the aims and methods of the agent. Sometimes an agent is doing a more calculative bit of reasoning with clearly defined rules and sometimes an agent is trying to think about broader

²²Though, depending on the content of the inconsistent propositions, it may, all things considered, not be worth it to bother dealing with the inconsistency. If it is not causing any harm, then it is not worth the cognitive resources to root out or figure out which beliefs are true once it is.

²³Brown and Priest (2004) explore a similar type of idea, but they are developing a formal system to explain certain types of scientific reasoning.

issues involving many threads of thought. It is helpful to ask whether we can give an accurate story of how the agent came to the decision by appealing to a simple set of rules or not. Generally speaking, we have made a lot of progress on picking out particular ways of reasoning and explaining how they work, but we have not made much progress in discussing how agents choose among ways of reasoning or in how they integrate results from different methods.²⁴ I believe that when continuing work on this problem, and in trying to theorize about human reasoning generally, it will be more helpful to drop the assumption that people are unable to handle an inconsistent belief set. I have started to show ways that this can be productive above.

I have argued that there are situations in which the rational thing to do is not only to believe inconsistent propositions, but also reason with all of them together. And these are situations that people do commonly find themselves in. I have argued this to bring out what is valuable and what is not in the fragmentationist view in order to better understand an aspect of human reasoning. In the next chapter I will dive into evidence-responsiveness and rational belief formation.

²⁴See Gigerenzer et al. 1999.

CHAPTER 3

Incomparability and inconsistency

The fragmentation view, as we saw in the last chapter, is one way of responding to two prominent problems that arise in conjunction with the possibility of a rational agent having inconsistent beliefs. These can helpfully be split into problems with how a rational agent could be in an inconsistent belief state on the one hand and how they would function once they got there on the other. In the last chapter I focused on the latter issue, raising doubts about the fragmentationist way of addressing it and sketching an alternative way of explaining how agents can be expected to reason and make decisions when their antecedent belief states are both inconsistent and not quarantined in the way that the fragmentationist suggests. But I have not yet said much about the issue of how agents get to the point where they have inconsistent beliefs in the first place. This issue actually splits in two. First, there is the basic problem of what sort of thing a belief is, and whether it makes sense for us to attribute inconsistent beliefs to any agent.¹ But even if we grant that an agent could end up with inconsistent beliefs, there is a secondary question of whether an agent could do so without running afoul of some rule(s) of rationality. In this chapter I tackle this last problem, setting aside the issue of attributing inconsistent beliefs full stop until the next chapter.

I have already started making progress on this problem by discussing the etiologies of the inconsistent beliefs in some core cases. In this chapter I will start by going over one interesting way of trying to ground the existence of strong consistency norms on rational belief: by deriving them from responsiveness to evidence. Being responsive to evidence, and

¹Which, in an interpretationist framework, is the same question as: is it possible for any agent to end up having inconsistent beliefs? We will return to this in the next chapter.

so accurately representing what the world is in fact like, is a core part of what beliefs are supposed to do. One of the constitutive functions of beliefs is that they serve as accurate representations of what the world is like, and so a well-functioning agent should have beliefs that are responsive to their evidence (evidence being how we get information about the world). If this argument grounding consistency norms in evidential norms works, it is a serious blow to the position that in some cases a rational agent can have inconsistent beliefs, given the importance of evidence-responsiveness to rationality. For the bulk of this chapter, I will directly deal with this issue of responding to one's evidence, only zooming out to discuss rationality more broadly in the concluding section. I will draw out an assumption in this argument and discuss how in some cases it is not warranted. Specifically, I will be discussing comparability of evidential support and in what cases it will fail. I will conclude that this reason for thinking that a rational agent must be consistent do not apply across the board. This means that it can helpfully illustrate what we find wrong with some types of cases of inconsistent belief, but they do not provide insuperable reasons that any rational agent must have a consistent belief state. We will also see hints that in fact the pressure to be consistent and the pressure to take one's evidence seriously may pull apart from one another. It may be the case that in order to be most responsive to our evidence we need to sacrifice a bit of consistency.

I will start by reviewing the purported connection between responsiveness to evidence and rationality. The argument I am focused on is due to Kolodny. One important aspect of his approach worth highlighting at the outset is that the purported irrationality attached to inconsistency does not hinge on the agent realizing that there is an inconsistency and thus realizing that there is something wrong. The error in rationality is present earlier, so it actually does not matter whether or not the person knows they are inconsistent; if there is an inconsistency, then one of the beliefs was erroneously formed. So, the mistake in rationality does not depend on the existence of a conflicting belief. This is one reason why this argument is a good stepping-off point to talk about the forming of inconsistent beliefs.

As a contrast, if the irrationality of inconsistent beliefs comes from the agent noticing the inconsistency and not responding appropriately, there cannot be a problem with forming inconsistent beliefs in the first place. The problem will only come in later as a failure to notice that one's belief set is inconsistent or as a failure to respond appropriately to the problem of having an inconsistent belief set.

After discussing Kolodny's argument and before turning to my own, I will warm up by reviewing another objection to the connection between evidence-responsiveness and consistency due to Alex Worsnip and looking at an alternate way of thinking of the structure of norms of rationality which appears separately in papers by Daniel Fogal and by Jim Pryor. My argument will begin with pinpointing the comparability assumption at work in the Kolodny argument. I will argue that not only does Kolodny not give sufficient argument for this principle, but it is likely false in some of the same interesting cases that we discussed in the last chapter. I will close out by summarizing what sort of conflict there may be between evidence-responsiveness and inconsistency more broadly, outside of the specific cases that Kolodny is concerned with.

3.1 Consistency and evidential norms

The argument I will be discussing in this section is a part of an ongoing debate about the relationship between substantive and structural norms of rationality.² Substantive norms have to do with responding correctly to evidence or reasons for belief. Generally speaking, they are about making sure that you are responding correctly to your situation or environment. Structural norms on the other hand have to do with relationships among an agent's beliefs. Consistency norms are structural and norms about how to respond to evidence are substantive. Worsnip (2015) frames the debate by discussing two ways that people try to collapse these two norms. Coherentists and some Bayesians try to reduce substantive norms

²See Fogal (2019).

to structural norms and Kolodny and some others have tried to reduce structural norms to substantial norms.³ Kolodny is doing this because he finds structural norms somewhat mysterious and thinks that their applicability to the rational evaluation of beliefs is less straightforward than the applicability of substantive norms.⁴

Kolodny's view on consistency norms is that they do not have an independently binding force. This is because of the reasons outlined in the previous paragraph; consistency norms do not have any clear connection to the primary function of belief. Because of this he wants to give an error theory to help explain why consistency norms seem to be binding which includes an argument that anyone who believes both a proposition and its negation is in fact violating a substantive norm of rationality and not just a separate (structural) consistency norm. More specifically, any agent who has inconsistent beliefs is, according to Kolodny, not responding correctly to their evidence. While I share some of Kolodny's attitudes towards consistency norms, I take issue with his error-theory for them. I believe that his argument about the relationship between evidence responsiveness and consistency does not go through, but it is connected to many important and interesting issues related to when it is acceptable to form a belief and thus serves as a good jumping-off point.

Let us start with the same structural consistency norm that Kolodny does, that one should not believe two propositions of the form $\{P, \text{not-}P\}$. He takes as a starting point the intuition that if someone believes both a proposition and its negation at the same time, then they are violating some requirement of rationality (232).^{5,6} He attempts to explain this

³Errol Lord (2014) gives a similar argument, but cashed out in terms of rationality and reasons. He explains means-end and closure incoherence with the idea of, roughly, reasons-transference. See also Raz (2005).

⁴The argument of Kolodny's that I discuss in this section is from "How does coherence matter?" (2007a).

⁵I will often, in the remainder of this chapter, drop the "at the same time" in the interest of not being repetitive. When discussing belief change over time I will make that clear.

⁶He extends his account to other ways of being inconsistent, like believing all of $\{P, \text{if } P \text{ then } Q, \text{not } Q\}$. The differences among the cases he gives an error theory for will not matter much for my argument, but I will be discussing cases where the inconsistency only arises when looking at a much larger set of beliefs later in this chapter. In these cases, Kolodny does not think that we can give a simple error theory on the model

intuition by appealing only to the rational requirement to be responsive to your evidence and starts by making this requirement both more strong and more precise. This version of the requirement (which he calls “strong evidentialism”) states that one is permitted to believe a particular proposition only if the evidence available to one supports that proposition more than its negation. He gives two motivations for strong evidentialism.

The first motivation is simply that strong evidentialism is plausible if one is already on board with evidentialism, which he states as follows: “There is reason for one to believe p only in so far as the evidence indicates, or makes it likely, that p is true, and there is reason for one not to believe p only in so far as the evidence indicates that p is false.”(233)⁷ He takes it that evidence has an obvious connection to rational belief, where consistency does not. This is because beliefs are, constitutively, about accurately representing the world. And our evidence is our way of getting information about the world. This is unlike avoiding inconsistencies, which is not directly tied to what the world is like because in many cases one could make one’s beliefs consistent by getting rid of accurate beliefs and keeping inaccurate ones. (Of course, there is some connection; if you have two beliefs that are not consistent with one another you know that they cannot both be true.) This makes evidentialism plausible, because one’s beliefs should be whatever is determined by what information we have about the world. From this statement of evidentialism and the idea that whenever the evidence an agent has access to indicates that P is true, then the evidence supports P more than not P , strong evidentialism follows. He does not argue for the position that whenever the evidence an agent has access to indicates that P is true, then the evidence supports P more than not P , but this is the most direct way to get from evidentialism to strong evidentialism, which he indicates is an easy jump. This particular way of filling out what it means for evidence

of the one for the ban against $\{P, \text{not } P\}$.

⁷It seems as though he should have said “disbelieve P ” instead of “not to believe P ” here. Later on this same page he states that one should not believe P if the evidence for and against it are evenly balanced, which is certainly not a case in which the evidence indicated that it is false. The first part of this principle is sufficient for my argument.

to indicate that something is true would also explain an intuition Kolodny has, and thinks is common, which is that if the evidence one has is equally strong for P and not P, one must have neither belief.

The second motivation is via a decision-theoretic analogue, where we have utilities associated with believing P when P is true and not believing P when it is false, belief states as actions and possible worlds as outcomes (234-5). If we assume that believing that P when it is true is less valuable than not believing P when it is false and that strengths of evidential support are probabilities (and that one can run such a decision-theoretic analogue in the first place), then strong evidentialism is the best strategy to follow to maximize utilities.⁸ It is important for understanding this argument that here he is assuming all-or-nothing belief and is not discussing credences. The probabilistic structure here attaches to how strongly the evidence supports the belief, and not to how strongly the agent believes the proposition.

If we grant strong evidentialism, the conclusion that one is not permitted to believe both a proposition and its negation follows fairly directly. It cannot be the case that a body of evidence supports P more than not P and also supports not P more than P for any proposition P, so one could never be permitted to have two beliefs where one is the negation of the other. So, if strong evidentialism holds, then we have an alternate explanation for why we feel that people who have inconsistent beliefs are violating a norm of rationality. (And so, we have a type of argument that a rational agent could never get into the situation where they have inconsistent beliefs of this type.)

So, how will this view work when applied to cases? Let us start with a quick word on the Carlos case from the first chapter. Kolodny and Davidson will both agree that he is irrational in forming the belief that he will pass the driving test, since he lacks sufficient evidence that he will pass. So, we can see that in having the inconsistent belief set including

⁸To say that the strengths of evidential support are probabilities is to say that they are numerical values that follow the probability axioms and the theorems thereof. E.g., they are real numbers between 0 and 1, etc.

both the proposition that he will pass the driving test and the proposition that he will fail, he is not substantively rational. He is not paying enough attention to his evidence. Other classic cases from the fragmentation literature also play well with Kolodny's story. When the agent is still fragmented it will seem to them when they are considering one fragment of evidence that they have sufficient evidence to believe P but not not-P and vice versa for the other set of evidence. But once they are looking at all of it together, they will be able to see that one of the beliefs is to be preferred. For example, the train navigation example works this way. Sometimes it will seem to Andy that he has sufficient evidence that Charing Cross road runs East-West and sometimes it seem as though he has sufficient evidence that it runs North-South. But if he came to think of both things together, he would realize there was a problem and (possibly after doing some more research) update his beliefs. This is also the case for other types of examples. Sayeqa is absent-minded because she is nervous about her history test later. Someone tells her that the school cafeteria is out of oatmeal. She forgets this and then reasons that there must be oatmeal left because they usually do not run out this early. If she forms both of these beliefs, the rational problem with her reasoning is that she did not take proper account of the testimonial evidence that she was presented with, not the bare fact that her beliefs were contradictory. (Though, if she notices that she had both beliefs, this could tip her off to the fact that something has gone wrong.) This is not how the cases I am interested in will turn out, as we will see later.

3.2 Rational pressures and conflicts

It will be helpful, before I get to the meat of my argument, to see some details of an alternate way of approaching the relationship between consistency norms and the norm of evidence-responsiveness. While Kolodny (and others, as noted previously) think that these different norms ultimately collapse in some way, alternatively one could think that they are separate

norms that could well pull apart.⁹ I will quickly review a few such positions. Ultimately I agree with the approaches canvassed in this section, but if Kolodny is right they will never come into play, at least in cases of a belief set of the form $\langle P, \text{not}P \rangle$, so it is still necessary to address Kolodny's argument.

3.2.1 Fogal on primacy of pressure

Fogal (2019) introduces a distinction among theories of how to evaluate the rationality of agents that will be useful for our purposes. One can have a requirements-based theory or a pressure-based theory.¹⁰ A requirements-based theory will hold that evaluating the rationality of an agent is a matter of seeing whether they meet or do not meet elements of a set of requirements. On this picture we can say that an agent is or is not means-end coherent, is or is not logically consistent, etc. Alternatively, we can evaluate the rationality of an agent by seeing how well they manage pressures of different magnitudes and directions. On this second view a rational agent should balance the pressures of, say, being logically consistent, with other pressures like being means-end coherent, intending to do what you believe you should do, etc. He takes it that this view is already common when it comes to substantive norms of rationality but is arguing for it when it comes to structural norms of rationality.

Fogal claims that the pressure-based view enjoys more intuitive support than the requirements-based account. He argues for this with the use of colloquial examples. One

⁹This is ignoring for the moment (along with other authors who discuss Kolodny's position on this matter) the difference between different consistency norms. Kolodny argues that it is true of several consistency norms that there is no way that they can conflict with evidential norms. But he does not think that his argument can be extended to a couple of other consistency norms, including, importantly for my purposes, the norm that would tell you that there is something wrong with preface paradox cases. We will get to this point in the next section.

¹⁰While Fogal is the best place to look for an explanation of the distinction, looking at rational requirements in a pressure-based way is certainly not new. Daniel Dennett seems to advocate for a pressures view (1987, 96). And Thomas Nickles (2002) makes a similar claim about how scientists operate. He claims that heuristic fertility trumps consistency and treats consistency as one rational pressure.

he discusses is a juror hearing evidence from the prosecution and from the defense in a criminal trial. He thinks that it is natural to speak of the effect of each of these as being a pressure, which at different times can mean that the person should be leaning one way or the other, or that they are agnostic. The pressure-based account also has the advantage of not being as vulnerable to counterexamples as the requirements-based account. We can see this by looking at how each account would handle an agent that seemed to be rational while not following some rule of rationality. The requirements-based account would have to find some way of explaining away the counterexample, by saying either that the agent is in fact not rational or that they did not actually break with the requirement. The pressure-based account could say that the person was in fact being rational while not following the rule because there was some over-riding pressure to break the rule. One consequence of this type of view is that, when judging the rationality of an agent, one will need more holistic information about their mental states in order to judge whether they are rational or not. This means that if all we know about someone is that they have an inconsistent belief set, we cannot judge them to be irrational. This is good news for me. But it is also worth noting that now the task of judging rationality has gotten much more complicated, as we cannot judge the rationality of a small number of mental states in isolation. I will now turn to an explanation of the preface paradox due to Pryor in order to turn to a more concrete example of this type of treatment.

3.2.2 Pryor on weighing norms

Pryor (2018) treats consistency as one norm of epistemic rationality among many. He uses a pressure-based view of rational requirements as part of a family of solutions to many paradoxes and puzzles of rationality. I find the solution to the preface paradox to be particularly relevant for the broader question of the prevalence and rationality of inconsistent belief states. In the situation of the classic preface paradox we have an author, Wilder, who

has written a book.¹¹ In the preface he includes a note that the errors in the book are his fault and not the fault of the reviewers. This implies that there are errors in the book. If we suppose that Wilder believes all of the claims that he has written in the book, including those in the preface, we find ourselves confronted with a paradox. Is Wilder being rational or irrational? On the one hand, he seems to be rational because he was epistemically responsible in forming all of the individual beliefs. The beliefs that he states in the body of the book were well-formed according to the rules of the scientific enterprise Wilder was engaged in (in this case mathematics). The belief in the preface of the book was formed on the basis of a justified induction. Wilder is a professional and is aware that (almost) all previous books to his that were published had some mistake or other, and he has no reason to think that his is any different. On the other hand, he seems to be irrational because his beliefs are inconsistent. Pryor responds to this seeming paradox by saying that Wilder is rational in the all-things-considered sense but irrational in the sense that his epistemic state does exhibit some normative defect, in this case being inconsistent. In Fogal's language, Pryor is using a pressures-based view to argue that Wilder is rational because while there is a pressure to be consistent, in this case that pressure does not override the initial reasons that Wilder had for forming the individual beliefs. Pryor narrates how we should think of Wilder's belief state in terms of the least-bad option available to him. Just accepting the inconsistent set is less bad than any of the alternatives available. Suspending belief in all of the claims of the book is not tenable (I will have more to say on this in a later section). Neither is picking some claim at random to drop from the set of beliefs expressed in the body of the book, as the choice would be entirely unmotivated. (Not to mention the fact that this will not get rid of the paradox. Wilder will continue to believe that some claim in the book is false, which will now correspond to a slightly different negation of conjuncts. So his belief set will still be inconsistent.) It would also be worse to drop the skeptical belief expressed in the preface, since he has good reason to think that it is true, not to mention that it is also connected to

¹¹See Makinson (1965).

the fact that he thinks that certain types of inductions are warranted, that previous books had certain mistakes, etc. Getting rid of that belief would cause a large change in his overall belief state that would be unwarranted.¹²

I would like to note that there is a structural similarity between the preface paradox and a more general paradox of rationality. Even those that do not write books think that we make mistakes, and one can see how a similar sort of paradox can get going with this bit of information. Among my beliefs is the belief that some of my beliefs are false. I will return to this sort of case later in the chapter.¹³

We can see the main problem of the current chapter at play in this example. In the preface paradox we can see a case where being responsive to evidence and being consistent can be thought of as different pressures of rationality that need to be weighed. Here it would seem that in being responsive to his evidence Wilder is committed to being inconsistent and if he were to make his belief set consistent he would need to ignore some of his evidence. In the next section I will talk a bit about how this idea can be applied to the case where an agent has the type of inconsistent belief set primarily discussed by Kolodny.

3.2.3 Worsnip and Kolodny on inter-level consistency

Worsnip (2015) directly argues against the idea that evidence-responsiveness and consistency will go hand in hand. This is in response to the two ways that people have argued that they collapse into one another. Recall that he goes over two ways that people can explicitly or implicitly connect these two epistemic norms. First, one can reduce evidence-responsiveness to consistency. This is the way that a coherentist about rational belief will go. Second, one can claim that consistency is reducible to evidence-responsiveness. This is the way that

¹²This is similar to the argument I gave in the previous chapter regarding the options available to the agent “resolving against the evidence.”

¹³Easwaran and Fitelson (2015) set up the preface paradox in this way (calling it the “global preface paradox case”), and also note that it is a case where consistency and responsiveness to evidence pull apart. Cf., Foley 1993 p 186 as well.

Kolodny and others with a similar position argue. Worsnip argues that these positions are both mistaken via instances of failures of KK, which demonstrate that there are cases where the best way to be responsive to your evidence is in fact to be inconsistent. KK is the principle that if an agent knows a proposition, then they also know that they know that proposition. He is able to construct these examples because in some classic cases of the failure of KK, the agent's evidence supports a particular proposition above some threshold, but does not support the proposition that the agent knows the proposition at that same level. But this means that the agent violates a particular inter-level consistency constraint that Worsnip argues should be considered to be a standard consistency norm. This constraint is that when an agent believes that her evidence supports a propositional attitude she should have the propositional attitude and when she believes her evidence does not support it she should not.

He gives the example of the proposition: "Charles I ascended to the throne in 1625" and a particular anxious student's epistemic relationship to this proposition. Sayeqa believes this and is justified in her belief and it is true, so she knows it. But she is nervous and does not remember her justification properly, so she does not know that she knows it. She has the evidence that it is true, but not the evidence that she has the evidence that it is true (Worsnip 2015, 21). So, in order to be responsive to her evidence, she should believe "Charles I ascended to the throne in 1625," but she should not believe "I know that Charles I ascended to the throne in 1625." But, if we also think that she should follow the inter-level consistency requirement outlined above, then she is also required to not have the belief that "Charles I ascended to the throne in 1625," since she believes that her evidence does not support that belief. So, Worsnip argues, she can either meet the rational requirement to be responsive to her evidence or she can be iteratively consistent, but she cannot meet both. So, consistency and evidence-responsiveness can pull apart. I will return to this idea of Worsnip's at the end of this chapter.

In light of this we need to stop and take stock of how far Kolodny's argument purports to extend. Does Kolodny's argument give us reason to think that Fogal, Pryor and Worsnip must be wrong about different norms of rationality pulling apart? Here is one way we could fill out an argument for that conclusion on Kolodny's behalf:

1. All rational norms are either substantive or structural. (Premise)
2. All substantial norms are evidence-based. (Premise)
3. All structural norms are consistency norms. (Premise)
4. Any rational pressures an agent could be subjected to must be either evidential or consistency-based. (1, 2, 3)
5. In order for an agent to be experiencing rational pressures in two different directions, one of three things must be happening: the evidence is pulling them in two different directions, there are two different consistency norms that cannot both be satisfied, or there is a consistency norm that is pulling against the evidence. (4)
6. Cases where an agent has some evidence for one conclusion and some evidence against that conclusion are not cases where there are norms of rationality that are pulling against one another. (Premise)
7. If an agent is responding correctly to their evidence, they will also be consistent. (Premise)
8. Different consistency norms cannot pull apart, and they cannot pull against the evidence. (7)
9. There are no cases where there are norms of rationality that exert pressure in different directions. (5, 6, 8)

The first step is a premise that all parties would agree to. The second step is implied by Kolodny's evidentialism. The third step is very in line with the way that Worsnip and Fogal seem to be carving up the conceptual space. Worsnip (2015) writes as though all rational norms will be evidential or about consistency. Fogal explicitly uses language about consistency when he introduces what a structural norm of rationality is (2019). Kolodny (2007a) writes as though there are other structural norms besides consistency norms, but the ones he discusses could arguably be included under a more broad umbrella of consistency

norms. One example is the principle he calls “believed reason.” Believed reason is very similar to the inter-level consistency norm that Worsnip discusses. Believed reason states that one should have the propositional attitudes that one believes their reasons require. It also states that one must suspend attitudes when one is entertaining a “live doubt” about them. Turning to the cases, on Kolodny’s view an agent approaches their body of evidence as a whole. We can either compare how strongly the evidence supports various propositions or we cannot. If we cannot, then the evidence as a whole requires us to withhold judgment. If we can, then our evidence clearly supports one over the other, and it is misleading to talk in terms of conflicting norms.¹⁴ Steps seven and eight follow fairly directly if we allow that Kolodny’s argument will extend to other consistency norms.

Unfortunately for this purported argument, Kolodny’s argument does not extend to other such norms, because step seven does not hold (even according to Kolodny). We can see that by looking at a toy preface paradox case. Let us say that the book itself has one hundred declarative sentences, and the agent writing the book has evidence that supports each of those sentences nine times as much as the evidence supports their negation. Then, supposing that the strengths of evidence behave probabilistically as explored above, the evidence would support the proposition “One of the sentences in the body of the book is false” more than its negation. In more detail, assuming that the sentences are probabilistically independent, then the degree of evidential support for the proposition that all of them are true is $.9^{100}$, which is approximately .00003. Of course, the actual number may go up a bit because the sentences are not likely to be independent, but it is clear that the number will remain well below .5, especially as the case becomes more realistic and more sentences are considered. So, Kolodny’s argument has not succeeded in completely collapsing consistency and evidential norms. But I will still need to discuss what is wrong with Kolodny’s argument if I want to claim that consistency norms as a whole can be one rational pressure among many in order

¹⁴If you do not find this bit of the argument convincing, I am not too concerned. The type of thing that Fogal, Worsnip, Pryor and I are arguing for is not that different bits of evidence push in different directions. That is a very uncontroversial conclusion.

to deal with the case, that I think is common, of a person believing both a proposition and its negation.

3.3 Comparability and the threshold for belief

In this section I will present my main argument. We have seen Kolodny's argument that responsiveness to evidence will imply consistency. In 3.3.1 I will start by arguing that Kolodny is relying on an unstated premise: that how much an agent's body of evidence supports a proposition is comparable to how much that body of evidence supports its negation (at least for any beliefs that the agent has). I will then motivate the idea that comparability of evidential support actually does not hold in all of the cases that Kolodny needs it to. This motivation will come from cases like those discussed in the last chapter. But, even if we can have rational belief without comparability of strengths of evidential support, it may still be the case for some other reason that evidence responsiveness and consistency travel together. I will close up the chapter by arguing that this is not true and build on the work in the previous section to say that these are in fact rational pressures that need not push in the same direction.

3.3.1 Kolodny's reliance on comparability

One important assumption working in the background of the Kolodny argument is that evidential support is commensurable or comparable. I will roughly follow Ruth Chang's use of the terms in her 1997b, even though her discussion of the phenomenon, along with most other discussion in the reasoning literature, is focused on incommensurability or incomparability of practical values (and not of what we are concerned with in this chapter: reasons, justifications for beliefs, or evidential support). In my usage, two things are commensurable if they can be measured on a single scale, for example by saying that w amount of x is twice as good as y amount of z . Two things are comparable but incommensurable if they cannot be numerically

compared in a way similar to the above, but they can be compared. Two things, A and B can be compared if one of the following three statements is true: A is greater than B, B is greater than A, A and B are equally great.¹⁵ If we have commensurability in an area we will also have comparability, because if we have enough structure to be able to measure things on a common scale, we will certainly be able to compare any two things.

Recall that Kolodny's argument that one can never rationally believe both a proposition and its negation relies on strong evidentialism, the thesis that one is permitted to believe a proposition only if the evidence supports that proposition more than its negation. Kolodny will have to say that in all cases where it is neither the case that the evidence supports P more than not P nor the reverse, the agent must neither believe P nor believe not-P. So, conversely, for any rational belief it must be the case that the agent can compare the strength of the evidential support for it against the strength of the evidential support for its negation. So, while Kolodny does not need to think that comparability of strength of evidential support holds for all propositions when paired with their negations, he does need to think that it holds often enough for all of our rational beliefs to exist. And Kolodny does not give us an argument that comparability between strengths of evidential support is something that will hold that often. We can see that by looking at the two arguments that he gives in support of strong evidentialism.¹⁶

The first argument Kolodny gives is simply that strong evidentialism will be plausible if we are already on board with evidentialism. But the original statement of evidentialism does not include anything that would indicate some sort of comparability consideration. As

¹⁵For reference, this is equivalent to comparability but not commensurability plus the trichotomy thesis in Chang's vocabulary. The trichotomy thesis is that there are only three comparability relations that two things can enter into: greater than (A, B), equal to (A, B) and lesser than(A, B). She adds the trichotomy thesis to leave open the possibility of comparability, but with more options for the comparison available. This will not matter for our purposes because I will be arguing that Kolodny relies on Comparability with the trichotomy thesis.

¹⁶He does briefly mention a third argument, or at least motivation, for this, which is that strong evidentialism would also explain an intuition Kolodny has and thinks is common, which is that if the evidence one has is equally strong for P and not P, one must have neither belief (Kolodny 2007, 233).

a reminder, he states evidentialism as follows: “There is a reason for one to believe p only in so far as the evidence indicates, or makes it likely, that p is true, and there is reason for one not to believe p only in so far as the evidence indicates that p is false.”¹⁷ In order for it to be the case that whenever the evidence makes it likely that P is true then the evidence supports P more than its negation, you need to build in extra assumption(s). One plausible way of doing this would be to assume comparability. For if we know that we can compare the evidential support for P and its negation (so either one is greater than the other or they are equal), and we know that the evidence makes it likely that P is true, then it would make sense to say that it supports P more than the negation of P , because we would not want to say that according to the evidence it seems likely that P is true if the evidence supported both equally or supported the negation of P more than P itself. If you recall our intuitive way of filling out this argument for him in section 2, comparability is built in; it included the assumption that in order for evidence to indicate that something is true, it must support it more than its negation. This is indeed an extra assumption, because while theories of what counts as evidence vary quite a bit, none of them imply commensurability or comparability of evidential support directly.¹⁸

This brings us to his second argument. Recall that it relies on the idea that we can treat strengths of evidential support as probabilities. This of course only makes sense if strengths of evidential support are not just comparable but are commensurable on top of that. It is interesting that this is the most filled-out argument Kolodny gives for strong evidentialism, because it is indicative that comparability (in this case the even stronger commensurability) is the most natural thing to reach for in order to justify strong evidentialism. He does not give an argument for the idea that strengths of evidential support will be precise in this way. He simply suggests that while one could object to the idea that strengths of evidence are probabilities perhaps the most natural response to this objection is to concede that

¹⁷See fn 7 for a comment on this formulation.

¹⁸See Rysiew (2011) for a broad-ranging discussion of what counts as evidence.

the agent may not be aware of them to this degree of precision. But from a God's-eye-view, they are actually probabilities. This does have one important implication from the agent's perspective. If the agent does not know what the precise relationship is between the strengths of evidential support for P and not-P, then the agent cannot come to believe either proposition on that basis. Adding some other assumptions that Kolodny takes to be plausible, this means that the agent must have neither belief.¹⁹ So, the situation is the same as far as the agent is concerned whether the strengths of evidential support are in fact incomparable or if they are comparable but the agent does not know which comparison relation holds.

Let us pause here to take stock. Kolodny thinks that strengths of evidential support are probabilities, though he has not given us an argument for why this is the case. He has allowed that an agent may well not be in a position to know how evidential strengths compare, even if they in fact do. But in order for his argument to go through, it simply needs to be the case that for every belief held by a rational agent, the evidential support for that belief is comparable to the evidential support for its negation. So, the question we need to examine more carefully is if there are cases where an agent can rationally believe a proposition even though the strengths of evidential support are not so precise (and are in fact incomparable). We will turn to this question in the next sub-section.

3.3.2 Etiology and belief formation

We saw in the last subsection that Kolodny needs comparability in order for his argument to go through. I will now motivate the idea that *incomparability* is to be expected, at least in the kinds of cases I started discussing in the last chapter. Also, importantly, we cannot simply suspend belief in these cases.

¹⁹The extra assumptions will not matter for our purposes, but they are as follows. 1) The agent is using either maximin or maximizing expected value on the assumption of equal probability (both classic decision theory strategies when one does not have enough information). 2) We should care more about avoiding false beliefs than having true ones (Kolodny 2007, 234ff).

There are two types of cases which will be particularly helpful here. Let us start with the resolution case from the last chapter. Someone is trying to quit smoking and they have one set of reasons to think that they will stop smoking, namely that they have decided to do so and they have control over their actions. They also have another set of reasons of a very different type to think that they will not, namely a simple induction on past attempts to quit. The agent's body of evidence will contain both sets of reasons. In light of this we have no reason to expect that we can compare the how strongly the evidence supports the proposition that the agent will quit smoking to how strongly the evidence supports the negation of that proposition.

For those that are uneasy on relying on these types of cases where the agent is predicting their own behavior, we can get a very similar case going that avoids this issue by taking the case of deciding what to believe about a loved ones promises and resolutions. Similarly in this case, one has gathering evidence impersonally as well as something like trusting that people will do what they say that they will do. I discussed a case like this at the end of section 2.2. Sometimes people are in a position where they have to weigh testimonial against observational evidence. Figuring out how to weigh these considerations is rather complex in the case where it is in service of deciding whether a person will do something in the future. Typically, if I have observed something myself I will not rely on the testimony of others to gather evidence about it, but in the case of an agent saying what they will do in the future their testimony carries a particular weight. In some cases, one of these may clearly outweigh the other, but generally speaking it will be a complicated matter to try to weigh these considerations. And again, we have no reason to think that in general agents will be able to compare these two sorts of evidence.

In the previous chapter I argued that there are strong reasons to believe each of these contradictory propositions. Here I am simply pointing out that Kolodny has not given us an argument that we can compare how strongly the evidence supports each of these propositions. Given how different the evidence in support of the different types of beliefs are and the fact

that people often do not know how to deal with these different sorts of reasons, the burden of proof is on him. Of course, the bare fact that the types of evidence are very different does not in itself mean that a burden of proof has been shifted. The scientific and manifest image example is one where the types of evidence are very different, but there is no issue with how to compare them against one another. The type of evidence I have in favor of “This tabletop is completely solid” and the type of evidence I have in favor of “this tabletop is mostly empty space” are entirely different. But there is no sense in which they are incomparable. If by “completely solid” I mean that it will bear weight so I can safely write or eat dinner here, then the simple observational evidence and any evidence I have about chemistry and microphysics work will both fully support the proposition that “this tabletop is completely solid.” If I am considering whether the table top is mostly empty space in a more scientific way, comparing how much of the volume of the table is occupied by particles as opposed to empty space, then the everyday observational evidence I have simply does not bear on the question and so is clearly outweighed by my scientific knowledge. But there is no clear winner like this in the Marušić cases.

Something may strike you as strange in how I have been discussing these examples. When introducing them in chapter 2, I wrote in terms of reasons for belief and not in terms of evidence. I did this in order to mirror more normal ways of talking about these types of cases and also in order to mirror the way that I talk about them in the other parts of this dissertation. But the difference between these two ways of talking obscures a potential problem with my argument that I have not discussed yet. The problem is whether or not all of the considerations and reasons for belief I discuss in the examples from this sub-section count as evidence or if some of them do not. In particular one may have worries about the agential mode of reasoning, as it is not the type of thing normally discussed in terms of gathering evidence. If they do count as evidence, then my remarks above can stand as written. However, one could argue that only the more observational evidence counts as evidence proper, and so the principle of belief-formation that Kolodny presents clearly rules

in favor of believing that the person will not follow through on their resolution in these cases. There is no issue of comparability because all of the evidence is in fact on one side.

For the sake of argument, let us suppose for a moment that these considerations do not count as evidence. In this case I will object to Kolodny's story much earlier, back at the first statement of evidentialism. This is because the things that we are taking into consideration here are in fact things that should (epistemically) be taken into account when deciding on what to believe. Defending this claim against the hypothetical evidentialist who believes that these considerations are not evidence and thus should not be counted when deciding what to believe would take us too far afield. Kolodny does not argue for evidentialism in this paper except to say that it is more obvious how evidence bears on belief than how consistency bears on belief. He also says that he is using the term "evidence" in a broader sense than usual, without saying specifically what will and will not count as evidence. I have been interpreting him as meaning that he is referring to substantive reasons for belief and not after-the-fact reasons derived from how beliefs fit together.

Having these concrete cases illustrates that it is not trivial to have comparability of evidence. In cases where we can weigh the evidence on a common scale, we may be able to get it fairly easily, though I suspect that straightforward cases like these are less common than you might think. We will also have comparability in cases where one proposition has overwhelmingly more evidential support than its negation. And in these types of cases, at least a certain level of consistency should prevail. That is to say, while we may well have reason to allow inconsistencies of other types, an agent should at least not have inconsistencies of the form "P" and "not P." But in the hard dilemma-like cases discussed immediately above, we have no reason to suppose that the agent's evidence base will have strengths of support that are comparable. But this also helps us to see one possible reason that we do expect a certain level of consistency most of the time. Most of the time, at least when settling a particular issue in a bit of reasoning, we will not have more difficult evidence bases like this; most of the time our evidence will be more well-behaved in one of the ways mentioned above. And

in that case, the argument of Kolodny's will go through, and we can expect our agents to be consistent on most propositions at least, just from following their evidence.

In response to this, Kolodny could double-down on the idea that the different strengths of evidential support are comparable, but the agent is unaware of how they compare. There are two different questions to ask about comparability of evidence: whether there is a fact of the matter about which proposition is better supported by the evidence and whether the agent is in a position to know which one is better supported by the evidence. Given the way he responds to the objection to his decision-theoretic argument, Kolodny seems to be supposing that there is a fact of the matter about how evidence compares, not that this information will always be available to the agent. But he has stated that in cases where the rational agent does not know how the strengths of evidential support compare, they are still not allowed to form either belief (regardless of one of them being in fact better supported by the evidence). So, if I can show that an agent can rationally form a belief without knowledge of how the strengths of evidential support compare, then that is going to be a serious problem for Kolodny and so for this way of arguing that rational agents must be consistent. And it is certainly not going to be of any help in understanding how rational agents should operate if it turns out that there are precise yet inaccessible ways of describing the relationship between their evidence basis and the propositions that they are considering.

I agree that I have not proven objective incomparability of evidence. But we have an issue regarding how high we want the threshold of belief formation to be for a rational agent. I think it is too much to ask for the agent to suspend belief in all cases in which they are not in a position to say that they have evidence that is stronger for a proposition than its negation. In the cases I opened this subsection with, I have shown that the agents are not in such a strong epistemic position. Nevertheless, they need to have full beliefs in order to function practically in these cases.

Kolodny and I are both working within a framework of full belief. So, here at least, we do not need to get into complications like credences or other graded pro-attitudes. The issue is,

then, whether two full beliefs or a suspension of belief on the question at hand is the better response to this particular epistemic situation. I do not believe that not having a belief on the question of whether oneself or someone else will keep a promise or a resolution is a viable option. If one has no opinion on the matter of whether one will smoke again, then it is hard to see how it is rational to, for example, give away one's cigarettes. But consider a decision to promise someone else that you will not smoke. If you do not have the belief that you will not smoke, then it is irrational and even immoral to make such a promise.

One may try to say that the response to having these two full beliefs that contradict one another should be to sometimes reason with one of the beliefs and sometimes reason with the other. This will not work. There are two ways of filling out this idea. One would be a fragmentationist view. I argued against this way of thinking of these cases in the last chapter. A second way of making sense of this idea would be temporal changes. Perhaps sometimes the agent has one belief and sometimes the agent has the other belief. But the arguments against the fragmentationist account also give us reasons to think that this proposal will not work. The agent who is only considering one of the beliefs at a time will be ignoring some of their evidence. In addition, it is not clear how the agent can rationally decide to flip between the beliefs. Given that they have sufficient evidence for each of them individually, it would either be arbitrary which one of them that they had at a given time or they would have to decide which. Neither of these options is rational. Arbitrarily deciding to have one belief over the other would be unmotivated. One could decide to switch what one believed for practical ends, but that would not be epistemically rational. And there is no reason for a stable evidence basis to support different non-time-sensitive beliefs at different times, so there would be no epistemic reason to decide to flip between the two beliefs either.²⁰

Now we have an argument on the table that one way of collapsing evidential and consistency norms will not work. We cannot rely on the idea that the evidence available to

²⁰By including “non-time-sensitive” I mean to exclude cases of time-indexed beliefs or evidence. A stable evidence basis could well support my having the belief “Amber is in her office” Tuesday at noon, but not having that same belief Sunday at noon.

an agent will behave nicely enough to not support inconsistent beliefs. I will say a few last things about the Kolodny paper before the concluding section where I talk about the overall take-away I have regarding the relationship between rational belief-formation and consistency.

While I think that Kolodny is right about one source of the pull of consistency norms (and, obviously, I am sympathetic to his main agenda regarding the strangeness of consistency norms on their own), it is only a tiny part of the story. We can use this explanation of his in these cases without thinking that it will apply across the board. But I have an alternate way of explaining some of the discomfort with inconsistent belief sets: computational load. Having inconsistent beliefs makes reasoning harder and puts a general strain on the system. I will circle back to this issue in the conclusion, after we have the apparatus from the next chapter at our disposal, but I would like to point out a couple of important ways in which this differs from Kolodny's explanation here. First, it goes nicely with the rational pressure idea of Fogal's. It makes more sense to say that there is some pressure to be consistent because it is computationally more tractable than to say that because inconsistencies can cause reasoning to be more difficult they should be banned outright. Many things that we do not want to ban can make a bit of reasoning more computationally complicated: considering more actions as options, taking more evidence into account, double checking results when something is important, etc. Second, it is an issue that only comes in after the inconsistent beliefs have been formed. The problem I am discussing is that it is harder to reason the more inconsistencies are in a belief set that an agent is using, and this is an issue that will only arise once the inconsistency is already in place.

For me these are both positive attributes for a purported problem with inconsistencies. I want to say that inconsistent beliefs can exist in a rational agent, so I am interested in saying that people have some reason to be consistent, not that they absolutely must be consistent in order to be rational. Also, if beliefs can be formed in very different ways, then we should not expect that there is anything rational or otherwise preventing the formation

of beliefs that are inconsistent. So I believe it makes more sense for a purported problem with inconsistencies to be one that kicks in after the beliefs are in place.²¹

While I have focused on Kolodny in this chapter, Bayesians are engaged in a similar project. Bayesians are also trying to reduce consistency to a more basic rational requirement, but they are going about it differently than Kolodny. (And their arguments will apply to probabilistic consistency as a whole, whereas Kolodny was only able to collapse some consistency norms.) When updating on new evidence, there is an assumption that the relevant probabilities that are needed for Bayesian updating are well-defined, and so one can measure what effect the evidence has on the overall belief state. If one's credences are probabilistically coherent (which is the credence analog of consistency), and one updates on evidence appropriately as it comes up,²² then one's credences will remain probabilistically coherent.

While there are many different ways that Bayesians argue for this position, the one that is most relevant for my purposes is the Gradational accuracy argument, introduced by Joyce (1998).²³ The core of the argument is the claim, proved by Joyce, that if an agent violates the laws of probability in their credences, then their credences are necessarily less accurate than they could be, according to some reasonable constraints put on what we should say about the accuracy of credences. The reason this is the most relevant argument for my purposes is that is similar to Kolodny's argument; it ties consistency norms to accurately representing the world. Before I move on, I would like to note that while I have focused on a particular strand of Kolodny's, he is not the only one to rely on similar points in their arguments or assumptions about consistency of beliefs. I have cited people making similar

²¹Graham Priest(2006) makes the point that any rational reason for a belief can also be a reason for a belief that is not consistent some other beliefs.

²²For the Bayesian this means updating according to Bayes' Theorem, which follows from the axioms of probability. (Though, of course, the requirement of how to use Bayes' Theorem to update one's credences does not.)

²³This argument is nicely summarized by Hájek 2008.

moves in the literature above. In the situation that a rational agent is updating on the basis of their evidence they will have to know the conditional probabilities that relate the evidence to the proposition. This will also effect the credence that the agent has in not-P because it follows from the axioms that $\Pr(P)+\Pr(\text{not-P})=1$

3.4 Conclusion

In the previous section I gave an argument that Kolodny has not proven that evidence-responsiveness will imply consistency, not even in the case of an agent believing both a proposition and its negation. It is not clear from the fact that Kolodny's argument does not work alone, however, that there is no connection between evidence-responsiveness and consistency. But we have mentioned several types of cases in the course of this chapter that show that sometimes when an agent is best responsive to their evidence they will be inconsistent. We borrowed Worsnip's case of inter-level inconsistency, which will seemingly be a common occurrence given the ubiquity of not being confident in one's reasons. We also borrowed the preface paradox, re-visiting arguments from Pryor that the way to best respond to it is to be inconsistent. And we can borrow from Easwaran and Fitelson the idea that the preface paradox actually means that everyone will best respond to their evidence only if they are inconsistent. We have added to this list the resolution cases from the last section.

What Easwaran and Fitelson call the global preface paradox case demonstrates that more or less everyone will be in this situation of not being able to both be perfectly consistent and perfectly responsive to one's evidence. While the other cases seem to rely on what might be considered to be common human failings, like having not kept a resolution in the past or not being sure enough of your evidence, the best human agents should be in the global preface paradox situation. All that is required is that the agent has a varied enough set of justified beliefs for the agent to be rational in believing that one of their beliefs must be false. This is

a rational conclusion to come to because it is likely that at least one bit of evidence that one formed a belief on the basis of was misleading or one made some understandable mistake in reasoning. But this means that our agent will have an inconsistent belief set.²⁴

Of course, in this chapter as a whole I have been focusing on responsiveness to evidence. But as we saw on page 55, the arguments and examples throughout this chapter have been treating evidence and reasons for belief as roughly synonymous, so we already have the arguments in place to say that if one is responsive to one's reasons then one need not be consistent and in fact one may feel some pressure to be inconsistent. But considering the fact that believing true things and not believing false things is often taken to be a large part of what it means to be rational, there is one other issue that should be discussed. In particular, we still have the problem that an inconsistent belief set must contain false beliefs.²⁵ Of course, if the inconsistency itself is the problem, then the agent is not going to know which of the beliefs is false. So, one could argue that avoiding believing falsehoods is important enough that one should suspend beliefs until the inconsistency is resolved. I think that this is too quick to throw beliefs out, for exactly the same reasons that I do not agree with Kolodny about suspending beliefs in all cases that do not have comparable evidence bases, as outlined at the end of the previous section.²⁶ But, given the ubiquitous nature of inconsistent belief sets, it would be particularly disastrous to have that as a requirement of rationality. Global preface cases are one example of this, but failures of KK like the Worsnip case above also may be extremely common. So, I prefer an approach to rationality on which it is a matter of responding appropriately to different rational pressures. This will help avoid issues attached to having too high a bar for belief.

²⁴See Easwaran and Fitelson (2015).

²⁵That is assuming that dialethism is false. If dialethism holds and truth is the main value we care about, then there are some inconsistencies that are very straightforwardly rational to believe. If, for example, one reasons one's way to {The liar sentence is true, the liar sentence is not true}, then one should believe both propositions, because they are both true. See Priest (2006).

²⁶I will discuss this argument again on pages 71f in the context of belief attribution.

This discussion can motivate a different way of thinking of reasoning as a whole. Coming to have a belief is not a well-delineated event that occurs in a vacuum, and we cannot evaluate it a such. We have larger ways of thinking and life projects that extend over time and have complicated relationships to other things in our lives. Because of this, evaluating the rationality of an agent is a complicated and defeasible process and not a matter of clear-cut rules that an agent is either following or not. Having too strict of conditions on when an agent is allowed to form a belief or being too quick to say that beliefs should be dropped would mean that we were recommending agents to have unnecessarily austere mental resources which damaged their functioning as agents. In order to be a better reasoner in a holistic sense, it may make sense to have an inconsistent belief set of one type or another. I turn to these issues of holistic evaluation and charitable belief formation in the next chapter.

CHAPTER 4

Interpreting the inconsistent

I will now defend myself from the objection that the very nature of belief forecloses the possibility that an agent could be in an inconsistent belief state. I have mentioned this issue in previous chapters in the context of discussing fragmentationists and their desire to use a possible-worlds semantics for belief, but here I am going to try to address what I take to be a more intuitive worry. Briefly, that there is a prohibition against attributing inconsistent beliefs to a rational agent, so agents cannot have inconsistent beliefs. This means that I will be working with an interpretationist conception of belief for this chapter. I have two related reasons for doing so. First, the interpretationist conception seems to pose more of a problem for the idea that we can have inconsistent beliefs than other conceptions. If we look at the representationalist conception, for example, there does not seem to be any problem from the point of view of what a belief is for the possibility of inconsistent belief sets. There is no obvious reason why I cannot represent both P and the negation of P in my language of thought.¹ But we will see that when we are looking at beliefs from an interpretationist point of view there are difficulties with the possibility of an agent having an inconsistent belief set. My second reason for using an interpretationist framework for this chapter is that I think it is a good one in which to explore an under-analyzed aspect of human reasoning processes, which will help us to make sense of inconsistent belief sets. There is a complicated and interesting relationship between the causal history of beliefs and the justifications that we give for them. Focusing too much on the justificatory relationships or taking them to

¹This is just one example. There are of course other ways of thinking of the representationalist mind.

be representative of reasoning processes as a whole can obscure some more messy aspects of human reasoning that are essential for understanding how these processes work.

These messy aspects of human reasoning will come out as I discuss a purported problem with inconsistent beliefs. My response will utilize a distinction between two different and sometimes conflicting constraints that we want our belief attributions to satisfy.² On the one hand we want to treat beliefs as causally efficacious mental elements in an evolving reasoning process. This constraint will be more salient when understanding the history or etiology of someone's mental states is important or when we want to get into the more complex elements of someone's reasoning. On the other hand, we think of beliefs as contentful states that have justificatory relationships with other mental states as well as the person's actions. Consistency standards are more helpfully and felicitously applied in service of the second constraint. More generally speaking, in different contexts we will prioritize these two constraints differently. And so, the importance of adhering to consistency norms will also vary depending on context. Sometimes allowing an inconsistent state to be attributed will best serve our ends and need not threaten our assumption that the agent is a rational being.

The discussion will start with a type of belief state that I will call transparently inconsistent. These are states including beliefs such that 1) the agent is aware of all of them and 2) the agent is aware that they are inconsistent. The focus will be on Davidson's reasons for believing these types of states to be impossible, but there is significant overlap between what Davidson has to say on the subject and common colloquial reactions to inconsistent beliefs. My response to this argument will utilize the distinction among contexts of belief attribution mentioned above and explained in section 4.2. After some discussion of examples, I will argue that there are in fact contexts in which attributing inconsistent beliefs to someone is the proper thing to do. I will close by walking through how this discussion also makes sense of why people are so reticent to attribute inconsistent beliefs to someone.

²I will be focused on beliefs for most of this chapter, but many of the things I will say about belief attribution are also applicable to the attribution of other mental states.

4.1 Davidson on inconsistent beliefs

To get a better understanding of what Davidson's position is, I will start by going over some direct quotations of his. I will then explain why his overarching positions of radical interpretation, the principle of charity, and the coherence theory of justification do not directly imply a position as strong as the one he wants to take on the possibility of inconsistent beliefs. To wrap up the section I will fill out a Davidsonian argument for his position.

A belief that it is about to rain would lose much of its claim to be just *that* belief if it did not have some tendency to cause someone who had it and wanted to stay dry to take appropriate action, such as carrying an umbrella. Nor would a belief that it is about to rain plausibly be identified as such if someone who was thought to have that belief also believed if it rains it pours and did not believe it was about to pour. And so on: these obvious logical relations amongst beliefs; amongst beliefs, desires and intentions; between beliefs and the world, make beliefs the beliefs that they are; therefore they cannot in general lose these relations and remain the same beliefs. Such relations are *constitutive* of the propositional attitudes. (2004c, 196)

Nothing a person could say or do would count as good enough grounds for the attribution of a straightforwardly and obviously contradictory belief, just as nothing could sustain an interpretation of a sincerely and literally asserted sentence as a sentence that was true if and only if D was both bald and not bald, though the words uttered may have been "D is and is not bald" It is possible to believe each of two statements without believing the conjunction of the two. (2004a, 200)

Contradictory beliefs about passing a test must each belong to a vast and identical network of beliefs about tests and related matters if they are to be contradictory. Although they must belong to strongly overlapping territories, the contradictory

beliefs do not belong to the same territory; to erase the line between them would destroy one of the beliefs. I see no obvious reason to suppose one of the territories must be closed to consciousness... but it is clear that the agent cannot survey the whole without erasing the boundaries. (2004a, 211)³

In the first passage we get one constraint on belief attribution. We cannot attribute a set of mental states to an agent if we do not, at least in large measure, see that they have appropriate relations amongst themselves, like intending to take the means to your ends and believing the obvious consequences of your beliefs. This is a constraint on an attribution of a larger set of beliefs (and other mental states). It does not give us anything like a specific threshold for when the logical connections and causal chains are too incoherent, but it clearly states that there is such a threshold. This is a more holistic constraint on belief attribution because it is taking into account a large number of mental states when making a determination. Davidson's view is holistic, because he thinks that you cannot attribute meaning to a few sentences in isolation or attribute just a few mental states to someone, in attributing one meaningful mental state to someone you are attributing a lot more than that. In this chapter I will be concerned with holistic constraints on belief ascription, constraints that deal with a large number of mental states. I will also be concerned with the interconnectedness of people's mental lives.⁴

In the second passage we get an absolute rule against attributing one kind of single belief to an agent. Lastly, we see that he also thinks that we cannot attribute two obviously inconsistent beliefs to an agent; if an agent has an inconsistent belief set, they must be unaware of the inconsistency or not aware of all of the beliefs at the same time. So, he does not think that transparently inconsistent belief states are possible. I am sympathetic to the broader holistic constraints he outlines and the ban on attributing a self-contradictory

³This is in reference to the Carlos example discussed in chapter 1.

⁴I am using Fodor and Lepore's understanding of holism here. "Holistic properties are properties such that, if anything has them, then lots of other things must have them too." (Fodor and Lepore, 1992, 2)

belief,⁵ but I think he is wrong that we cannot attribute a transparently inconsistent belief set to an agent. And, in fact, I will argue that the broader holistic constraints and the rule against transparently inconsistent beliefs will come into conflict. In order to make sense of this, we have to separate Davidson's theories and overall project from what he says about the particular case of a transparently inconsistent belief set.

I will start by discussing radical interpretation and the principle of charity.⁶ At first glance, these related ideas would seem to be the part of Davidson's theoretical apparatus that would most directly tell against the possibility of inconsistent belief sets. Davidson uses restrictions imposed on us from the possibility of radical interpretation to deal with issues in semantics and epistemology. The details of these arguments need not concern us here, instead we will be focused on what this core theory of his has to say about what beliefs must be like. Radical interpretation is the interpretation of the utterances of a speaker without any prior knowledge of the language they are speaking or what their mental states are. This sort of thing must be possible because the only way we have of knowing what someone means by an utterance is by knowing what their mental states are, and the only way we have of knowing what their mental states are is by interpreting their utterances. Of course, most of the time we are not in this dire of a position. However, according to Davidson, we could not have gotten into the position of knowing what a language means in the first place if radical interpretation were not possible. Also, even when we are interpreting what someone is saying in a more normal conversational context, there are often holes that get filled in and mistakes that get corrected based on assumptions about the speaker: what they are likely trying to say, what they might believe, etc. Davidson breaks into the circle of needing to know what someone is thinking before knowing what they are saying and vice versa by proposing the principle of charity. We perform both tasks at the same time: interpreting

⁵To be more precise on the second point, I am not necessarily convinced that we cannot attribute a self-contradictory belief to an agent, but I will not be taking a hard stance on this issue in the current chapter. I will content myself with pointing out the costs of sticking to this restriction.

⁶See esp. Davidson (1984a).

what the utterances mean and attributing beliefs to the speaker. We accomplish this by assuming that the speaker is a rational agent similar enough to us. We can assume that they have perceptual faculties similar to ours, that they form largely true beliefs about the world, that they make simple logical inferences, that their beliefs are largely consistent, and that they mean to say mostly true and relevant things. This will allow us to interpret their beliefs and the meanings of their utterances. Without these assumptions, Davidson thinks, interpretation as a whole would be impossible.

As it turns out, the possibility of radical interpretation itself does not foreclose the possibility of inconsistent belief sets (even if the agent is aware that the set is inconsistent). This is because all that we need in order to get radical interpretation up and running is that people's utterances and beliefs are true and consistent *for the most part*. We can see this in the first block quote above, and also in "A coherence theory of truth and knowledge" where he says "but there is no chance that a person's beliefs will not tend to be consistent, since beliefs are individuated in part by their logical properties; what is not largely consistent with many other beliefs cannot be identified as a belief" (2001, 155). All of the statements of what he thinks on this particular issue are hedged like this. I do not have any problem with the idea that there is a pressure towards beliefs being consistent. And I agree that there will need to be some level of consistency within an agent, though trying to quantify where the thresholds of rationality are here is extremely difficult. The coherence theory of justification is similarly not by itself why he will have a problem with attributing an inconsistent belief set to an agent. (I will briefly touch on the coherence theory of justification later in the chapter.) This is because someone having some inconsistencies in their belief system does not conflict with the person's beliefs being largely true. But having inconsistencies in the beliefs puts a "strain" on the attribution of a set of beliefs to an agent (2004d, 217). Nothing in what he has said here implies an absolute ban against attributing a single contradictory belief or an inconsistent belief set that an agent is aware of, so to get where our disagreement lies, we will have to fill in some intuitive arguments for him.

One complication immediately comes up for this view, and should be clarified before we move on. What does it mean to say that someone's beliefs are "largely consistent" or "consistent for the most part"? I will be talking about levels of consistency throughout the chapter. More direct inconsistencies will count more than less direct ones. E.g. having a proposition and its negation both in a set will be worse than having the negation of a complex tautology in there or an inconsistency involving a large number of propositions. Other things being equal, a set that cannot be made to be consistent without removing a large number of beliefs is less consistent than one that would be consistent if only one or a small number of propositions were removed. The way that I am wielding this concept in this chapter does not demand that we have a way of comparing sets pairwise to see which one is more consistent. The issue is whether the inconsistency is bad enough that it cannot count as a belief set had by an agent. Davidson and I will have different sorts of answers to this question and neither of us have a precise answer.

Davidson himself makes it clear that the possibility of radical interpretation does not directly imply that we cannot attribute an obviously inconsistent belief set to an agent. Near the end of "Incoherence and irrationality" (2004a) he says that there is a line dividing types of seemingly irrational belief states. On the one side are sets of beliefs that we can attribute with difficulty to others while still being able to assume that they are a rational being with holistically interpretable mental states. On the other side are sets of beliefs such that once we attribute them to another being, we can no longer claim to be treating them as a rational agent with interpretable mental states; we are breaking the attribution of a set of beliefs to an agent rather than just putting a strain on it. (And, to be more accurate, in this case they are not beliefs at all.) This is why I said earlier that the beliefs are *seemingly* irrational. At this point you cannot say that they are irrational because they are no longer the sorts of things that can be judged as rational or irrational. They are arational, so making this sort of judgment would be a category mistake. Davidson draws the line so that attributing two beliefs that contradict one another but the agent is unaware of having both

is on the strained-but-possible side of the line and attributing two beliefs that contradict one another and the agent is aware of having both is on the impossible side of the line. I think that we can attribute beliefs under more strain. I will be spending the rest of the chapter arguing that the Davidsonian way of drawing this line is too limited and making sense of an alternate picture.

Many people will want to draw the line where Davidson does. What follows is my best reconstruction of an intuitive line of thought that is similar to Davidson's way of thinking about this issue and is also similar to things that I have heard from others in discussion and debate on this topic. I will start with an argument about a single contradictory belief and then talk about how to extend it to inconsistent belief sets.

1. A rational agent will not believe something they know to be false. (Premise)
2. Rational agents know that contradictions are not true. (Premise)⁷
3. A rational agent will not believe something they know to be a contradiction. (1, 2)
4. Rational agents are capable of recognizing simple contradictions of the form "P and not P." (Premise)
5. A rational agent cannot believe a simple contradiction. (3, 4)

The first premise of this argument follows from the nature of belief. The rest of the argument fills out why the idea of a contradictory belief might seem very paradoxical: people cannot be said to believe something that they know is false and they should be in a position to know that a contradictory belief is false.

At this point we only have an argument that we should not attribute a single contradictory belief to an agent. But I believe that many people's reasons for not wanting to attribute an inconsistent belief set to an agent are similar. There are a couple of ways that one could

⁷I will not be discussing dialetheism in this chapter. People giving this sort of argument will typically not be dialetheists. But someone could be a dialetheist about some small category of beliefs and still think that this sort of argument goes through just fine for most types of beliefs.

extend this line of reasoning to apply to sets of beliefs that are inconsistent. One is closure under entailment. There are different versions of closure under entailment, but even a fairly conservative one will get this result. For example: If an agent believes all of $\{B_1, \dots, B_n\}$, is aware of all of the beliefs, and is aware that $\{B_1, \dots, B_n\}$, entails C, then the agent believes C. If we allow an agent to believe an inconsistent set of beliefs and also think that their beliefs should be closed, then they will also believe a single contradictory belief, which we have already disallowed. So, agents cannot have inconsistent belief sets. Even if one does not want to endorse full closure, it can seem very strange to people that an agent could believe two propositions and not a simple conjunction of them (as long as they were aware of both of them at the same time). So, it makes sense to say that an agent at least cannot have two beliefs of the form “P” and “not P.”

Another way of extending this way of thinking into an argument against the possibility of inconsistent belief sets is to apply the connection between truth and belief to the set of beliefs instead of to the individual beliefs. To take the simplest case, let us suppose that an agent has two beliefs where one is the negation of the other. In this case, the agent is aware that they have both of these beliefs, and the agent is aware that together they are inconsistent. The agent then is aware that at least one of their beliefs must be false. And it seems reasonable to think that if an agent has a set of beliefs and they know that one of the beliefs is false (but they do not know which one), the agent should suspend the beliefs until they find out more. I agree that this is how it will work in many cases, but I take issue with this as a universally applicable rule regarding correct belief forming and updating processes.

Davidson explicitly gives reasons along these lines for the position he espouses in the second block quote above. He takes it that the only way we can evaluate someone’s rationality is to see if they are in line with their own values (Davidson 2004b); but we can assume, partially because of the principle of charity, that all people actually do value consistency (and in fact all of the values of classical logic and decision theory).

To sum up a version of Davidson's position here:

1. We need to be able to interpret the beliefs of others. (Premise)
2. We cannot interpret an agent at all if we do not assume that they are within the bounds of rationality. (Premise)
3. We cannot interpret an agent of having a set of beliefs such that the relationships between them are too far outside the bounds of rationality. (1, 2)
4. A transparently inconsistent belief set is too far outside the bounds of rationality. (Premise)
5. We cannot interpret an agent as having a transparently inconsistent belief set. (3, 4)
6. An agent can only have a given set of beliefs if we can interpret them as having that set of beliefs. (Premise)
7. An agent cannot have transparently inconsistent beliefs. (5, 6)

Lines 1, 2 and 6 are just part of the interpretationist conception of belief that I am assuming without argument for this chapter. I object to line 4. I agree that we should not attribute a belief to someone if we know that that person thinks that the belief is false. But I disagree that this means that we cannot attribute an inconsistent belief set to someone, even if the person knows that it is inconsistent. In this situation, where the agent knows that there is some false belief in a set, the interpretive situation is more complicated. In order to say that we could not charitably interpret the person as having the inconsistent beliefs, it will matter what other interpretive values are at stake and what other rational pressures the agent is under. I will be arguing that there are two important things that Davidson's account is missing. First, that there are a lot of different contexts in which we ascribe beliefs to an agent. Not all of them are such that it is theoretically desirable to demand such a high level of consistency. Second, that there are different rational pressures that we must be responsive to. It will not do to try to be charitable by avoiding attributing a transparently inconsistent belief set to someone and thereby accuse them of a worse epistemic sin.

4.2 Contexts of belief attribution

I will now start filling in my interpretationist picture. In what follows I will be almost exclusively talking about cases where an agent is making sense of why they believe something or did something. I have several reasons for this. Firstly, one of the central activities epistemologists and philosophers of mind are engaged in is describing in some sense or another why an agent's mental states are as they are. So, these contexts are particularly relevant for my project, which includes making sense of why philosophers have some of the assumptions outlined above about what kinds of belief states are possible. Secondly (and relatedly), these cases are also the ones that are more interesting from a theoretical perspective because they centrally involve more complex relationships among mental states. (This is in contrast to the types of cases often referenced in radical interpretation contexts, where the relationships that are highlighted are those among speech acts, beliefs and comparatively simple features of the external environment.) Thirdly, these are the contexts in which I think that an attribution of a particular kind of inconsistent belief structure is warranted. Near the end of the chapter, I will briefly discuss simpler cases and how they relate to my disagreement with Davidson. Most of this section is a discussion of some different trade-offs we make when we attribute beliefs to an agent in more detail, before we get to some examples in the next section.

As I mentioned earlier, Davidson's view is a form of interpretationism. So, for him mental states are most fundamentally things that we attribute to others in the course of interpreting their observable behavior. The constraints discussed above are not the only ones that he has on which sets of mental states we can attribute to someone. In addition to the principle of charity, he also has constraints on semantics, including compositionality and lawlike behavior of Tarski-style T-schemas.⁸ But, even under all of these constraints, his interpretationism means that there may be two different sets of mental states that one could equally well ascribe to an agent given the evidence one has. In this situation there

⁸See Davidson 1984b.

is no fact of the matter about which one obtains. Importantly, he does not think that this implies that mental states are less than real or vague or ambiguous. He compares different belief-attributions to expressing a temperature in imperial versus metric units.

Parts of my picture roughly line up with Davidson's here. I believe that there will often be more than one available interpretation of a person's mental states, including the interpreter's. Knowing what we believe, knowing why we believe it, and describing it to others is not a straightforward process of introspection. I would add that this type of interpretationism shares commonalities with some more general complications that arise when modeling a complex system. In this situation we will need to make simplifying assumptions. What assumptions those are, how much detail goes into the description of the system we end up adopting, what causal and rational aspects are brought to the forefront, etc. are all things that can vary across descriptive reconstructions. Two important structural features that we need to capture in order to be studying or understanding the mind of another are causal relations and justificational relations. Which of these things we will prioritize in our re-description and how much detail is tracked will vary by context. While I agree with Davidson that interpretationism does not make mental states any less real or substantial, I am interested in more substantive differences across interpretations than any properly analogized by a change in units. Alternative interpretations can change how we think about the person's mental life and what we are describing about them.⁹

The central trade-off is between etiological fidelity and justificatory structure, as introduced above. It will be helpful to start the explanation of this trade-off with a mathematical example. When one is trying to prove a mathematical theorem, one uses many different kinds of reasoning. One might look for matching terms in definitions, diagram geometrical situations, try out cases, etc. Much of the reasoning involved in figuring out if the mathematical fact is true and trying to construct a proof will not be directly reflected in the final

⁹Dennett, the other prominent interpretationist about belief, discusses the problem of having to discuss beliefs and desires in terms that are more precise than these mental states themselves satisfy (1987, 21).

version of the proof itself. So, there is a distinction to be drawn here between why the person believes the proposition in an etiological sense and how they would justify the truth of it to others within standard mathematical practice. This is not to say that the etiological story is arational or irrational. When one, for example, draws a quick back-of-the-envelope diagram to test a hypothesis, this can supply a good reason to form a belief or look into a particular way of proving a proposition.

While the above is much more regimented than a piece of everyday reasoning, there are structural similarities. Let us take an example of someone trying to decide how to vote on a local proposition to raise taxes on cigarettes. She starts by supporting it because smokers use more public funds on average than non-smokers. But then seeing an opinion piece about how parking tickets are essentially a way of unfairly taxing lower-income people causes her to go into a more serious period of discovery where she looks at facts about the average income level of smokers and other more on-the-ground facts. She then comes to oppose the proposition. If she is justifying herself to others, she will say something simple like “I am opposed because it is an instance of unfairly taxing people with less income.” If she is having a longer conversation with a friend who cannot make up his mind, she may tell a longer and more complicated story. So, while the rules of conversational justification are looser than the rules in standard mathematical practice, we can still see a rough version of the discovery/justification distinction. And, to go along with that, we can see a distinction between making the process of discovery clear to another person and making one’s current best justification clear to another person. There are one-sentence justifications, long and messy etiological re-constructions, and many points in between.

Another distinction that correlates with the justificatory/etiological distinction is the difference in level of ownership and responsibility that the agent has with respect to the relevant mental states. Beliefs expressed in a more justificatory context are usually taken to be endorsed in a full-throated way by the agent. Of course, in some contexts people can just try ideas on for size or play devil’s advocate. But it is more common for an agent to have a

little bit of distance from their own mental states if they are treating them in an etiological way. If someone is reporting their own reasoning history, they may be speaking of attitudes they could not control or attitudes that they no longer have.

Relatedly, the standards by which one would judge an agent's reasoning processes and offered justifications are very different, partially because while people have control over their actions and over what sorts of justifications they endorse, they do not have direct control over what they do and do not believe full stop. This is related to the common observation that one cannot "believe at will." When we are evaluating a justification that has been given for a particular belief or action, we can judge according to whatever standards of reasoning or behavior seem appropriate to the situation. Were the presented reasons in fact evidence for the claim? Did the person make a logical error? Did the person pursue the means to their ends? What we take proper reasoning to look like will affect what we put in these types of belief attributions. I will assume that people are behaving more or less according to what I take to be reasonable standards for behavior and reasoning. Similarly, when I reproduce my reasoning for someone else, I will want to bring the rational rules to the forefront in addition to possibly simplifying things for my audience. This can also occur when one is trying to make sense of one's own reasoning, make a hard decision, or decide what to believe about something. Think of the statement "I am scared of spiders because they have too many legs." The judgment of the rationality of the speaker will be different depending on what type of "because" is meant by them. If someone is offering number of legs as a legitimate justification for a fear response, there seems to be some sort of rational mistake. But, if the person is presenting it as a bare causal fact that they have no control over, the situation is different. The way in which we would judge the mistake is ameliorated.

Formalism will have a different role to play when one is primarily concerned with justification than when one is primarily concerned with etiology. There are a couple of important aspects of formalizations that are relevant here. First, they state things in a precise language. This part will not be important for my argument, but I believe that precise language

will be less useful for etiological redescription than it will be for justifications. Trying to describe mental causality is difficult and probably for the most part it is not useful to try to be extremely precise. Justifications are another matter. Second, they have rules of how different sentences in the language relate to one another. This can vary by formalism but can include things like rules for which sentences can be proven from which other sentences and which sentences must be true whenever some other sentences are true. Common formalisms applied to mental states include things like deductive logics, probability theory, default logics and decision theories.¹⁰ The nature of all of these formalisms mean that they are more helpfully and felicitously applied in contexts where the focus is on justification and not etiology, because this is what they were designed to do. Classical logic, to take one example, gives precise rules about truth relationships and provability. Those are both justificatory relationships, not causal ones. That one proposition entails another can give you a good reason to believe the latter but does not necessarily cause you to believe the latter. Of course I will have to give a more complicated story about what kinds of logical justifications are allowed. If we let inconsistent beliefs into the picture, an agent is not justified in believing just any classical consequence of their beliefs. Classical consequences within a mode of reasoning will be fine, but what kinds of justifications are allowed in a more general sense is a more complicated story.

Harman (2002) has made a similar point here. He claims that logic is not a theory of reasoning and should not be used as such. But logic is a theory of justification, since, obviously, it is useful to be able to show that the belief you are trying to justify is necessitated by some easier to accept premises. He makes this point in terms of induction and deduction, and claims that searching for an “inductive logic” is simply mistaken, because human reasoning is inductive but we should not think that a logic describes how human reasoning works in the first place. But we can take his point about the relationship between reasoning and

¹⁰Though some have explored the idea of paraconsistent probability theory. See Bueno-Soler and Carnielli (2016).

systems designed to formalize semantic and syntactic relationships without adhering to his ideas about the relationship between induction and logic. For example, you can think that logic is still a useful tool for justification and not reasoning while still thinking that there can be a good inductive logic.¹¹

Let us take a moment to consider how this issue intersects with the primary question at hand, that of the role of consistency norms in belief attribution. The working definition of an inconsistent belief set that we have been operating with relies on the idea of satisfiability, which is a concept from logic. Since justifications can naturally involve some logical structure or at least a formal syntax or semantics, it will be easier to say when a justification is inconsistent. Whether a set of beliefs is consistent is a helpful question to ask when one is trying to determine how good a justification is, since inconsistencies throw a wrench in most formal systems. A rational person will often take the fact that some belief set is inconsistent into account while they are reasoning, but the relationship is less straightforward.

When we are engaged in making sense of our beliefs and our actions, the level of detail at which we should and typically do describe the state of play as well as what kinds of relationships among the elements will be different depending on who we are talking to. When trying to make our inner workings more apparent to ourselves or those closest to us, it will often make sense to give longer stories that include more in the way of an etiology. There is a small group of people, including yourself, that has a particular interest in getting to the bottom of what makes you tick and random historical accidents that have been important to you. But when we are in a more public setting the expectation is that we are justifying ourselves to others or trying to convince others of something, so there is an assumption that we are making public states that we agree with and that we take ourselves to be justified in what we are saying. So, these different contexts will correlate with what kind of audience we have for the interpretation that we are offering up. It is also in keeping with rules of cooperative conversation that you present things according to simple and shared inference

¹¹For more discussion of this issue, see MacFarlane (2004) and Steinberger (2019).

rules. Things like modus ponens, generalization from cases, instrumental reasoning and other ways of making different mental states hang together in a rational way.

These distinctions are the building blocks of what I have to say about inconsistent belief attribution. In the next section I will dig into some examples before zooming back out for some takeaways from them and finishing up with some thoughts about what this distinction really means for the attribution of inconsistent beliefs, including why people resist attributing inconsistent beliefs.

4.3 Examples of belief attribution

In this section I will go through some examples that will highlight how different these two contexts of belief attribution are, as well as some assumptions that commonly go into belief attribution practices can make us miss things about reasoning processes. In the next section I will tie this thread together with the question more directly at hand: how these things can easily lead one to be wrong about the possibility of inconsistent belief.

I will start by discussing some complex actions, which will often occur over a longer period of time. What is important about these complex actions is that the agent is performing the action before she is aware that she is doing so. My first example is becoming someone's friend. You first start seeing them somewhere, say, at work. You notice you enjoy being around them, you start spending more time with them. Eventually you notice you have made a friend. There are times early in the development of the friendship when it would be true that you were making friends. But there is no way for you to know that that is what is happening yet, as other things can look identical but not be part of a budding friendship. Another similarly structured example is creating a piece of art.¹² You may make a sketch that you later use as a study for a painting. We would then want to say that the creation

¹²Thanks to A.J. Julius for this example.

of the painting started with the sketch, but at the time of the sketch you had no idea you would make a painting on that theme.

The important thing about these examples is that there was no decision before the beginning of the action. The reasoning process was more gradually evolving than that. This means that if we only pay attention to standard justificatory practice, in which we treat all actions performed by an agent as things that they decided to do and then did according to some set rules of practical reasoning, then we are missing important aspects of human mental life, in this case reasoning processes that are simply not studied enough. Many human reasoning processes do not have well-defined boundaries and evolve in ways that we cannot describe by keeping too close to standard justificatory relationships. If we keep ourselves to discussing means-end relationships, implication and others more standardly described in logical practice, we will not be able to say what is going on when someone is making a friend or engaging in an artistic project.

There are also cases that involve belief more centrally. Coming to have a belief or cease to have a belief is also a process, and it does not always have clear boundaries. Simple matters of fact you will tend to just stash away or quickly look into when you need to know them. But more complex evaluative beliefs (including beliefs about your own past), religious beliefs and beliefs related to learning a skill will not be so simply formed or unformed. For example, an someone coming to believe that cats make better pets than dogs do when they were previously a dog person. The process of changing beliefs in this situation will look similar to the process of making a friend. One starts spending more time with cats, learns more about them and has a lot of experiences with them them. The person does not sit down to figure out which pets are better, rather their evaluative beliefs change slowly over time.

It will also be interesting to take a look at a bit of reasoning that is closer to a case that would be typically discussed in terms of structure and timing. Take an example of someone choosing which university to go to. Suppose that Kristin is choosing between going to a state

school and going to a private school that is farther away and she chooses to go to the state school. Kristin may have at the time thought that it was clear that she chose to go to the state school because of the quality of its education program. It may later become clear to her that financial considerations were clearly the most important factor for her actual reasoning. We can imagine that she came to realize this through reflection and talking to others about their reasoning regarding these issues. In this case it seems like the considerations she explicitly thought about or would have given in response to questioning are not the same as the considerations that were most important to her choosing the school that she did. The unconscious motivating factors that were relevant to the choice that Kristin made are something that should play an important role in a theory of practical reasoning, because it is an important part of why people do the things that they do. This is one possible way that the agent could be confused their own reasoning. In cases like this we will not see a difference between etiology and justification in the immediate descriptions by the agent. But it is still a case that illustrates that we do not get all of the information that we need from given justifications. Other types of cases will illustrate this point in a more straightforward way. I believe that many cases are like the voting case above, which I will have more to say about later in this section.

One last type of case that I believe is worth discussing falls at the opposite end of the complexity spectrum from our friend case. That is, atomic actions, those that are extremely simple in the sense of having no actions that are constituent parts of themselves. In contrast, a complex (or molecular) action like making breakfast will have many smaller parts, like turning on the stove, cracking the eggs, etc. A great example an atomic action, due to David Velleman (2015), is standing up from a seated position. He uses this case to claim that reasoning does not necessarily precede action and sometimes only supervisory and perceptual thinking will precede an action (340). This supervisory thinking is the main element of his view. According to Velleman we are like passengers in a car: we pay attention to what we are doing and only intervene when necessary (344). He claims that we can only

argue for the view that practical reasoning is supervisory by an introspective experiment. While he only discusses atomic actions like those discussed in this paragraph, his language seems to imply that he is making a claim about practical reasoning more broadly speaking. The claim is that there is no reasoning that takes place before such an action, and that if we properly introspect this will be obvious to us. Instead, we just perceive that we are standing up, and only consciously intervene if necessary. While deliberation in a classic sense will sometimes precede actions, the fact that we speak about all actions this way is more of a matter of convenience than accuracy, and atomic actions are one example of this. Previously in this section we can see other examples of this. We do not, at least generally speaking, sit down and deliberate about whom to befriend. And things that affect our decisions may not be transparent to us. Smaller things that are not really worthy of sitting down and deliberating about are another example of this sort of thing. I do not deliberate about whether to turn on my kitchen light to make breakfast in the morning. It is a little misleading to even say that I do it for simple instrumental reasons. I do it more out of habit than anything.

Why does this lack of deliberation matter? It ties in to my broader points about the assumptions we make when we are attributing beliefs to others or making our reasons clear to others. Saying that I turned on the kitchen light so that I could see so that I could boil water so that I could make tea so that I could get a caffeine boost and get to work is right in one sense but misleading in another. It is right that we can attribute these mental states to me and they do largely reflect something, albeit simplified, about my reasoning processes. It is misleading in that I did not stop to do that sort of deliberation and even if somewhere along the motivational chain there was a missing reason I would likely have done the same thing. It is a common occurrence for people to make small steps towards things they are in the habit of doing even if they do not want to do them on that particular day. I may turn on the kitchen light and start boiling water before I realize that I am going to a breakfast where coffee will be served.

A pattern I would like to call attention to here is solidifying beliefs for the purposes of deciding what to do. This is why we spent so much time on practical reasoning in this section. This pattern of having a complicated ongoing reasoning process and then solidifying it for a moment for practical purposes is clearer to see in the practical case. But our belief states need not necessarily stay in the solidified state after the action has occurred. This can include verbal actions such as giving arguments and justifications to others. And most of the time when we are performing the action of explaining our mental states it is in a context that is focused on justification and not etiology. This is tied to the fact that we cannot assume that people are willing to listen to our life story whenever we are trying to express our beliefs. I want to head off a possible source of confusion here. We certainly do not want to say that there should be no relationship between a justification given to someone else for an action and an etiological reconstruction. If someone gives a reason for action that is unrelated to their reasoning processes, then they are lying. It is hard to say how much of a relationship is required, but the details of this will not be relevant for my argument here. You may only pick out a subset of your motivations to formalize, or what makes up your mind is really insufficient and then you fill out a proper justification later.¹³

We can see this pattern with the voting case discussed in the last section. The agent has prior beliefs about fairness that are not entirely well-defined or precise. In order to either explain her position to others or to check a box on her ballot, she needs to precisify what she thinks. But she may well have to do some of that work over again in order to use those beliefs for future practical questions. It is easier to judge her rationality based on how she decided or explained herself in a particular situation than it is to judge on the broader and fuzzier network of beliefs as a whole. Another way to see this difficulty is to think about the

¹³Of course, we are deeply social animals, so sometimes rehearsals very similar to what we would do with others can be done when we are by ourselves. Mercier and Sperber argue in *The enigma of reason* (2017) that human reason has two main functions corresponding to the two main challenges of human interaction (208). The first is the attribution of reasons which serves a justificatory function. The second is reasoning which serves primarily an argumentative function. They emphasize the point that I am agreeing with here, which is that people produce reasons for things they believe or do when in fact there either were no real reasons to speak of or the causal factors were in fact different.

strangeness in asking someone why they are friends with another person. Typically, when that type of question is asked the implication is that something more specific should be done, e.g. stopping the friendship at this point.

So, here we have two different ways in which we attribute beliefs to another person and some interesting relationships between the two. We have hinted at some ways in which this is more broadly interesting, but now I will switch back to my response to Davidson, which will utilize a look from a different perspective at some of the examples from previous chapters.

4.4 Attributing inconsistencies

As previously stated, I disagree with Davidson that a transparently inconsistent belief set is too far outside the bounds of rationality, and so his objection to the possibility of such belief sets does not go through. In a Davidsonian framework the possibility of transparently inconsistent belief sets revolves around their rationality, so my reasons for disagreeing with Davidson here are connected to my arguments from the previous chapter that it can be rational to have an inconsistent belief set. In this section I will briefly revisit those and with the help of the contexts of belief attribution apparatus from this chapter, show how they are relevant to Davidson in particular.

4.4.1 Rationality again

The problem I have with the idea that transparently inconsistent belief sets are too far outside the bounds of rationality is that if we set that as a hard and fast rule, we will end up having to be less charitable overall. If we insist on never attributing an inconsistent belief, it forces us to be less charitable because people have other rational pressures that they should be responsive to, a la Fogal and Pryor from the previous chapter. In particular, if we want our theoretical reasoning to be tracking evidence, not ignoring reasons for belief, and giving us an understanding of the world, we can better do that without strict consistency rules.

But, in addition, if we do not allow the occasional inconsistency in a belief set that we are attributing to another agent, then we will also be forced to cut other beliefs that make the set less cohesive or hang together less well as a whole. I will flesh out both of these ideas here. First, I will briefly revisit my reasons for taking evidence-responsiveness to be more fundamental to rationality than consistency. But recall that the reason we have from Davidson to avoid attributing inconsistencies is based on the idea that they are the paradigm of rationality and connectivity among a set of beliefs. So secondly, in case the Davidsonian is not impressed by the first argument, I will talk about ways that disallowing transparently inconsistent belief sets can cause problems for other connections among beliefs of an agent.

First, I fully agree with the previously stated reasons from Kolodny to think that responding to our evidence and reasons is more important than being consistent. Beliefs are, constitutively, about representing the world and guiding action. Evidence, broadly construed, is how we get information about the world. Being consistent is not as directly tied to what the world is like because in many cases one could make one's beliefs consistent by getting rid of accurate beliefs and keeping inaccurate ones. Recall that in the previous chapter evidence was used very broadly to include reasons like resolutions and agential powers. So, all I am claiming here is that substantive reasons for belief should be weighed more heavily than after-the-fact consistency norms. And we have already argued that evidence-responsiveness and consistency can pull in different directions than one another in the previous chapter, so we have one reason to think that Davidson is wrong here.

You may be worried that I am rushing past an argument of Davidson's for the opposite conclusion. After all, we discussed in the last chapter the fact that some coherentists are interested in reducing evidence-responsiveness to coherence. And while Davidson is not the most paradigmatic example of a coherentist about justification, he does think that coherence yields truth. Here is a reconstruction of his argument from "A Coherence Theory of Truth and Knowledge" (2001):

1. Given how beliefs are formed, in order for a radical interpreter to understand the meaning of the utterances of an alien, they must assume that the things the alien is saying are mostly the way the world is. (Premise)
2. Anything your beliefs can mean that could be understood by a radical interpreter must for the most part be the way the world is. (1)
3. For a belief to correspond with reality, the meaning of the belief must be the way the world is. (Premise, from definition)
4. Any part of your beliefs that could be understood by a radical interpreter must mostly correspond with reality. (2, 3)
5. There is no meaning to beliefs beyond that understood by a radical interpreter. (Premise)
6. Mostly beliefs correspond with reality. (4, 5)
7. Coherence of beliefs yields correspondence. (6)
8. Correspondence is the test of truth. (Premise)
9. If our beliefs are for the most part coherent, we have good reason to believe them to be true. (7, 9)

There are two things to note about this argument. First, there is no mention of evidence specifically. Davidson is here not interested in discussing evidence or particular reasons for belief. Also, Davidson is not mentioned by Worsnip (2015) when discussing the coherentist position, which makes sense as Davidson is not the paradigm example of a coherentist, and later on claimed that the position espoused in this paper is not best understood as a coherentist one.¹⁴

But, granted all of that, there still may be an argument here that undermines my position. Fortunately, I am safe from it, for the same reasons that Davidson's arguments that beliefs must be consistent for the most part are not a problem for me. All of the conclusions about what beliefs must be like that we see here are also qualified with a "mostly," either implicitly or explicitly. So, even if we tied evidence directly to reality, correspondence, or truth, we

¹⁴He discusses this in his afterthoughts on "A coherence theory of truth and knowledge," see 2001 p 155, quoted above.

would still not have an argument here that there are no cases where evidence and coherence pull apart.

This brings us to a more important point. Davidson is very focused on consistency values as the bedrock of rationality.¹⁵ But if we are committed to not allowing transparently inconsistent belief sets, we need to sacrifice other connections (that possibly fall under the banner of coherence relations) that Davidson will want to preserve. This will play out differently in different cases.

I have gone over how people can arrive at having inconsistent belief sets in different situations in earlier chapters, so here I will briefly recap and then emphasize a point about this that is relevant to someone who emphasizes connections among beliefs. In the resolution case we have reasons to think that we will follow through and reasons to think that we will not. If either the belief that the agent will follow through or the belief that they will not is missing, then there are negative epistemic consequences. Previously I talked about things like not being able to make decisions properly or not responding to evidence properly. But there is another way of looking at the problem. As Davidson notes, part of what it is to have a belief is for the agent to make simple inferences on the basis of it. We have these two ways of thinking about human agents, introduced by the promising or resolving against the evidence cases. On the one hand we have the agential mode of reasoning. This takes seriously the idea that people are rational and have a will that they execute and will tend to do what they want to do and what they say they will do. Within this mode of reasoning there are connections between the belief, for example, that my friend will stop smoking and my beliefs about agents that led me to have the belief. I will then go on to make inferences on the basis of this belief. I agree with Davidson that this is part of what it is to have these beliefs. On the other hand, we have a more third personal mode, where we take impersonal observational evidence more seriously. There are similar connections between beliefs associated with that mode of reasoning.

¹⁵See especially his 2004c.

When it comes to preface paradox cases we have a similar problem. If we suspend the skeptical belief, then that is severing an inference that should ideally be made, it is disrupting a kind of connection among the beliefs. There does not seem to be a very good reason why we are not allowed to make the inference in this case, but we are in other structurally identical cases. Continuing to apply this consistency rule would result in a very large number of beliefs being suspended. The case can be even more dire if we try to turn to the regular scientific beliefs instead. And it is still worse if we turn to the global preface case. We are then in a position of either allowing a bit of inconsistency or of demanding that agents arbitrarily cut relationships among beliefs or suspend staggeringly large numbers of them.

So, we see that we can get more connectivity among beliefs if we allow the occasional inconsistency. I have been ignoring the “transparent” part of the description of the belief sets, but this is because I argued in chapter 2 that it will not do to say that the agent should not have access to all of the beliefs at once either, so we will have more coherent belief sets that make more sense holistically if we allow small bits of inconsistency.

4.4.2 Belief attribution again

Before moving on to some concluding remarks, I will summarize what this has to do with belief attribution specifically and why it might still seem like a mistake to attribute a transparently inconsistent belief set to an agent.

To clarify, I still think that *most* of the time we do want to avoid attributing inconsistent beliefs to another agent, or ourselves. Most of the time we are operating in the justificatory context of belief attribution. And when we are just trying to make our best reasons for thinking or doing something clear to another person, we should avoid confusing the issue with complex etiological information. When interpreting the beliefs of others in that context, we should similarly assume a simple and consistent justificatory structure connecting their mental states. Given that this is how we operate most of the time, it is unsurprising that people have a strong aversion to attributing transparently inconsistent beliefs to someone.

Another thing to back up my point here is the choice of examples. In the stereotypical radical interpretation case, it is obvious that we should not be interpreting a speaker as communicating inconsistent beliefs. In a setting like that, the ideas being conveyed will be comparatively simple and relevant to everyone involved in the conversation. If I am communicating with someone from the ground up, I will be saying things like “that is a rabbit,” not trying to convey my complex inner life. In a case like that if it seems as though if I think I have heard the speaker utter something of the form “P” and something of the form “not P,” I should think that I misheard or misinterpreted something, or possibly that they have changed their mind. If I am having a more casual conversation with someone and they utter both a sentence and its negation, I will most likely interpret them as being forgetful, changing their mind, or as not really meaning both of the sentences and rather using that as shorthand for having mixed feelings or something of that nature. But if I am more familiar with someone who has been wrestling with an issue and they communicate what seems to me to be a transparently inconsistent belief set, I may take them at their word on what they believe. The same should go doubly if I am studying a case of reasoning about a difficult matter as a theorist.

This is related to an issue that we glossed over previously, which is the cardinality of the transparently inconsistent set. We should expect stronger connections to more important persisting beliefs in the case where the inconsistency can be shown with a smaller set. So, we need more motivation for the set {I will stop smoking, I will not stop smoking} than for a set in the style of a preface paradox. The latter will be less of a reason to be concerned and will also complicate one’s life less. In the preface paradox we are necessarily dealing with a large set of beliefs. There is not an inconsistency otherwise. In the resolving against the evidence case we are also dealing with a large set of beliefs, but in this case it is not just the small set containing the two contradictory beliefs, but rather the contradictory pair and the larger backdrop of beliefs connected to those two beliefs.

As I mentioned earlier, I am in agreement with Davidson about the broader interpretationist framework, but I want a little more flexibility at our disposal to make sense of a holistic web of mental states. We can hold onto the idea that when attributing mental states to another we should be charitable and able to handle some level of radical interpretation while occasionally allowing for transparently inconsistent belief states. And in fact, if we allow such states, we will be better able to make sense of longer-standing connections among the larger set of mental states.

The fact that we do have different ways of attributing beliefs to an agent is a good thing. Having the ability to focus on messier etiological reconstructions or on neater justificatory relations gives us a more robust and detailed picture of how people's reasoning processes work in general, as well as a more full picture of the mental states of any particular agent. Multiple models are a better guide to understanding. It will likely turn out that different imperfect models are a better guide to understanding something than one model that is missing important pieces in order to avoid being wrong about anything. We can see this by looking at the example above. The ways we have of making a consistent, justificatory model of the agent's behavior are missing important elements. So, in this case I am recommending that the philosopher thinking about the mental lives of agents takes a similar tack to the scientist modeling an ecological system.¹⁶

I am sympathetic to the idea that agents need temporary consistency of a sort in order to act. But even if there are cases where someone temporarily resolves a complex and possibly inconsistent set of beliefs for the sake of an action, this need not give us a reason to think that the person is consistent in general. The person deciding whether they are going to smoke again or not can decide that for the sake of making a costly bet with a friend they will behave as though they will not smoke again. This does not mean that the matter is settled and deciding to treat the case as someone changing their mind rapidly is unmotivated. Their

¹⁶See Matthewson and Weisberg (2009) for an argument that a single model cannot maximize all desirable qualities of a model.

reasons for having the inconsistent beliefs have not changed. They have just decided to act in a particular way in one particular situation.¹⁷

4.5 Conclusion

I have just argued that in fact we will make more progress in understanding reasoning if we sometimes attribute inconsistent beliefs to an agent. Further, this does not tell against many Davidsonian positions on philosophy of mind and epistemology. In this section I will address a couple of objections, and then have a brief word about what I think these cases mean for human reasoning more generally.

4.5.1 Objections and replies

In the previous section I made a case for why people feel compelled to not attribute inconsistencies to others. What I have not explained is why an agent would be disturbed to come to be aware of inconsistent beliefs in themselves. There are several explanations that we could give on the basis of what I have said so far in the dissertation. First, I would like to note that if this seems like a problem that presents itself to be solved and not a problem that instantly solves itself then that counts in my favor that transparently inconsistent belief sets are possible. Remember that for Davidson if you come to this realization, you can no longer be in the inconsistent belief state. So, if you come to the realization that you have inconsistent beliefs and then need to decide what to do about it, then that is evidence in favor of the possibility of having inconsistent beliefs.

But, setting aside the focus on possibility of this chapter, we need to say something about why people almost universally see inconsistent beliefs as a problem. (And, of course, what

¹⁷Priest gives several more straightforward examples of someone acting as though they have inconsistent beliefs. The one I find more plausible are: asserting contradictions, writing a book about dialethism, disagreeing with classical logic, etc. (Priest 1986 104).

people see as a problem is also relevant for how to most charitably interpret others.) First, remember that I am on board with saying that inconsistent beliefs are probably not fine if they are generated in the same way. The only situations in which I want to say that it can be all-things-considered rational to have inconsistent beliefs is if they were not all generated in the same mode of reasoning. They need to have some distance between them, as was discussed in chapter 2. But this does not explain the more general phenomenon. For that, we have to look to my alternative explanation for why people think that having inconsistent beliefs is a bad thing: that it makes reasoning more complicated and so indeed there is a rational pressure against having inconsistent belief states.¹⁸ While I do not think that people are explicitly thinking that it is a bad thing because it is more cognitively cumbersome, I do think that people have a feeling of not knowing what to do with inconsistent information, which tracks with what I am claiming. When we see that we have inconsistent beliefs, we will not know what to do with them as they are, they are only helpful to us with their background etiologies. Another thing to consider is that the prevalence of giving justificatory belief ascriptions will rub off on us in all sorts of contexts. We want to give consistent, justificatory stories and when we cannot we see that as a bad thing.¹⁹ So I do not think that there is one unified explanation for the discomfort that people feel in ascribing inconsistent beliefs to others. It is due to the focus on justificatory practice, the fact that in many cases an inconsistency is due to a simple mistake, and also that inconsistent information is harder to deal with.

There is an unpleasant dilemma facing anyone with the view I have been arguing for here. We need to either deny a very conservative form of closure or we need to say that people have beliefs of the form “P and not P.” Take the version of closure from earlier in the

¹⁸Dennett states that we should assume people are consistent at first, but that this assumption can be discharged (Dennett 1987, 21 and 45). He similarly wants to ascribe mental states to people using the principle of charity.

¹⁹This is similar to Dennett’s point that we are likely to ascribe more linguistic structure to mental states than they in fact have, given what tools we have at our disposal for theorizing about them, namely language (Dennett 1987, 21).

paper: If an agent believes all of $\{B_1, \dots, B_n\}$, is aware of all of the beliefs, and is aware that $\{B_1, \dots, B_n\}$, entails C, then the agent believes C. The very simple instance of this that we are considering is an agent believing they will smoke, believing they will not smoke, being aware of both of these beliefs and believing in adjunction. We then need to either say that they do not make that simple inference or that they believe “I will smoke again and I will not smoke again.” In fact, on the first branch of the dilemma we are not just denying closure, which would imply that the person would have to make that inference, but we are denying something even weaker, which is that they can make that inference. But, given that we already have to deny that people are justified in coming to believe the classical consequences of their beliefs without restriction (because of explosion), the first branch of the dilemma might look more appealing.

4.5.2 Concluding remarks

One last thing I would like to comment on before I move on to the conclusion is mental inertia. This is one thing that is well-captured by the etiological context of belief-ascription and not by the justificatory one. People do not start fresh when they are deciding what to do or what to believe. They have a large weight of things that they typically do and already believe and they tend to not deviate radically from that. Taking this and being able to communicate good, shorter justificatory bits to people is a good thing and helps us to make our thoughts clearer and change our minds about things. But if we think that people are starting from a place of trying to get the best-justified set of beliefs they can based on their evidence, we are missing an important part of the picture. Coming to realize that there is an inconsistency in your belief set will not mean that you immediately suspend them in all cases, because many are too well-entrenched for that. They have too many connections to other things that we believe, or want, or plan to do.

This is one thing that is structurally captured by Iris Murdoch’s moral psychology. She de-emphasizes the moment of decision in favor of the moral development that leads up to

it.²⁰ According to her, it is a struggle to see things accurately, and once you are there, what to do in a particular situation is clear and there is no longer a problem to solve: what to do or believe is clear if you understand the truth of the situation. While I think that her endpoint is idealized, possibly past the point of usefulness (though she does say that the final moral end-goal is not attainable), I do think that she is entirely correct about where the bulk of deliberative effort comes in. We know from studying human behavior that habits drive what people do. They can be broken, but through effort. We can interpret Murdoch as discussing habits in thinking and how to break them and make new and better ones: practicing looking at the broader world as opposed to yourself: what you pay attention to.

I think that she did a good job of emphasizing an important part of how people reason broadly, not just about moral issues. We reason using worldviews that we spend a long time slowly building up and changing. This alone does not imply that it is best for people to be inconsistent, but it is a helpful piece of the puzzle. If we have different modes of reasoning that we develop throughout our lives that are rich and different enough from one another, then inconsistencies are inevitable. And people's mental lives are too complex and valuable to say that the response here should always be to go back and suspend beliefs.

²⁰ “[A]t crucial moments of choice most of the business of choosing is already over” (1971, 37).

CHAPTER 5

Implications

In this conclusion I will address one lingering objection before saying a few things to wrap up some loose threads from previous chapters.

One proposal that I have argued against in the dissertation is that when one is confronted with an inconsistent set of beliefs the response should be to suspend enough beliefs to solve the problem. A related proposal that I have not addressed is that the agent should not have beliefs that are inconsistent, but it is alright for them to have inconsistent pro-attitudes of another kind. This could be because the agent did have inconsistent beliefs, but upon realizing that they suspended their beliefs and formed new states. Or it could be because the attitudes were never strong enough to count as beliefs in the first place.

Taking the discussion from the last chapter as a starting point, our interlocutor could argue that what we are doing in the cases where we have reason to attribute “inconsistent beliefs” to an agent is actually attributing an inconsistent set of some other mental state. Since I have talked about the attitudes being used in reasoning, both theoretical and practical, it will be useful for the sake of filling out our interlocutor’s position to use a distinction borrowed from Catherine Elgin. She sets up a contrast between belief and acceptance.¹ Belief is tied to truth and acceptance is tied to action. But it is important to note that the conception of “action” here is broad and includes theoretical actions, like reasoning. In fact, for her purposes she restricts the use of “acceptance” to attitudes that are for cognitive as opposed to practical purposes and also rules out extremely temporary cognitive uses, like

¹See Elgin (2017, 19). She is elaborating on a distinction due to Jonathan Cohen (1992).

things that are accepted for the purposes of a reductio. Using this distinction, we can characterize our interlocutor as saying that people can accept an inconsistent set of propositions, but they cannot believe it.^{2,3} We can see how that would make sense on Elgin’s way of carving up the conceptual space. The inconsistent belief set in my cases is something that is actively being used in reasoning by the agent. But there is some slight issue when it comes to the truth, because we know that the beliefs in an inconsistent set cannot all be true.

I do not object to surrendering the term “belief” to people who want to use it for states that have a tighter connection to truth than the states that I am talking about. But in this case, beliefs become less central to epistemology.⁴ This is especially apparent if we look at the generalized preface paradox case. If I try to describe an agent as having the belief that not all of their beliefs are true, then the interlocutor in this case would have to advocate for saying that the agent does not have any beliefs, and only has things that they accept. At best, they could maybe have a few beliefs here and there that they were extra certain about.

But, setting aside the extreme case, we have other reasons for thinking that if this is how we are dividing up types of mental states, we should be looking to acceptance if we want to study how people reason. The things that the agents accept in my examples are attitudes that are tied to evidence, represent what the world is like to the agent, and help to guide action. They are absolutely essential to figuring out how an agent is reasoning about the world day-to-day, and are different from the type of philosophy of science examples that Elgin uses, like ideal gas laws.

²Their reasons would be, presumably, similar to what we filled out in the last chapter on the part of the person who does not think that it is possible to have inconsistent beliefs.

³Rescher entertains a related idea, that we have beliefs separated into those that are more certain and those that we accept in a more guarded way (1988, 5.2).

⁴Elgin suggests something similar in her book, partially because epistemology is a normative enterprise and people do not have control over what they believe. More on that in a moment.

5.1 The structure of reasoning

I have chosen to press on the issue of inconsistency because I think that it is a good way of getting into broader structural features of human reasoning as a whole. I have argued that we can (rationally) form inconsistent belief sets and we can function just fine with them. This is because sometimes inconsistencies will come up as a natural result of the fact that we have different modes of thinking about the world that have their own separate inertial force and holistic structure. And we will do more harm in trying to disrupt this system than the harm that is done by these systems not agreeing with one another. Some of this is obscured by different ways we have of talking about beliefs. We are often, for good reason, focused on a justification for a particular belief or action. When analyzing a small bit of reasoning, it is often fine to prioritize things like consistency, but sometimes we need a broader picture. The researcher in the preface paradox case is a good example of this. After seeing corrections come out all the time for other books that are not essentially different from theirs, it would be hard not to infer that there will be an error in their book. This inference has the inertia of a lot of history behind it and it cannot be overruled without good reason. They have observed a lot of instances to support the induction they are making, they have a history of automatically making enumerative inductive inferences, etc. Something similar can be said about the first-order research side of things. They have a developed method for their field of inquiry.

An important part of this picture that we have seen over the past few chapters is several ways in which consistency plays an important role in reasoning. It is helpful for regulating within a mode of reasoning, it is helpful to make reasoning simpler, and it is helpful to make us organize our thoughts for others. (And these boons of consistency do not depend on a particular formalism.) With these purposes for consistency in mind we can also see a couple of types of situations in which it is helpful to point out an inconsistency, even if sometimes it is rational to believe inconsistently. It is helpful to point out if the inconsistency

is indicative of a mistake or if the person is in a context where consistency is expected. For example, pointing out an inconsistency in a philosophical argument will be helpful because, most of the time at least, philosophers are in the business of putting together a consistent justification or exploration of a particular position. So, if we looked at all of an individual's philosophical beliefs there may be inconsistencies, but it is still helpful to consistency check individual arguments, papers or positions. The same goes for many less formal arguments among friends. If someone tries to give an inconsistent justification while trying to convince me of something, it is more than appropriate to call them out on it. Inconsistencies do not belong in that context. But if someone is thinking out loud or just explaining why they think something they do, it may not be appropriate. Lastly, inconsistencies in many contexts are simply the result of a mistake made somewhere. If I am simply predicting someone's future behavior on the basis of their past behavior, I should not come to two incompatible conclusions. Or if I infer that Amber must be in her office but I get a message from her that she is in the library, I am making a mistake if I continue in the inconsistent state of believing that she is in her office, she is in the library and people cannot be in two places at once. I should update my belief state so that it is consistent.⁵

Depending on how they came about in the first place and how they will be used in the future, we will treat different inconsistencies differently. Except in the special cases that we have talked about here, where different elements of an inconsistent belief set come from different modes of reasoning and it is not clear how to prioritize between the two, inconsistencies are due to a mistake in reasoning. In general this sort of inconsistency should be dealt with by updating the belief set so that it is no longer inconsistent. (Sometimes inconsistencies can arise on matters that one will not think about again, in which case it is probably not worth the cognitive resources to sort out.) But even in the case where the inconsistent belief set itself is not due to a mistake, there is still some pressure to clean up

⁵This is really a long way of saying that in many cases, I agree with Kolody's view on how inconsistencies will not pop up if you are just reasoning correctly.

the beliefs, because reasoning on an inconsistent belief set is a complex matter; this means that in the short term one will often need to prioritize one over the other or create a sort of temporary consistency for the sake of a particular decision. So, as a result of this picture, we have a nice view on what is valuable about consistency.

We have a picture of reasoning here, but still a lot of questions about how to fill out the theory. I have here argued for shifting our way of thinking about agents towards their long-term management of different ways of reasoning about things and away from individual decisions or justifications for beliefs. But I have not said much about how specifically agents should be integrating different modes of reasoning. There is also a large open question of how to evaluate the rationality of agents. We have greatly complicated the story by saying that it is a matter of managing pressures in different directions. There are pressures to be consistent overall, pressures to keep connections solid within a mode of reasoning, pressures to respect very different sorts of evidence, etc. Since we have no hard and fast rules of rationality we have a very complicated story to tell about how to evaluate the rationality of large sets of beliefs. And we do not have much on the table to help us with this issue yet. I am skeptical about the possibility of formalizing this high-level balancing act that agents engage in. But we should know better than to condemn an agent for an inconsistency absent any other context.

Bibliography

- Aristotle 1984. *Metaphysics*. In Barnes, J., editor, *The Complete Works of Aristotle*. Princeton University Press, Princeton, NJ. Translated by W. D. Ross.
- Borgoni, C., Kindermann, D., and Onofri, A., editors 2021. *The Fragmented Mind*. Oxford University Press, Oxford, UK.
- Brown, B. and Priest, G. 2004. Chunk and permeate, a paraconsistent inference strategy. part i: The infinitesimal calculus. *Journal of Philosophical Logic*, 33(4).
- Bueno-Soler, J. and Carnielli, W. 2016. Paraconsistent probabilities: Consistency, contradictions and bayes' theorem. *Entropy*, 18.
- Chang, R., editor 1997a. *Incommensurability, Incomparability, and Practical Reason*. Harvard University Press, Cambridge, MA.
- Chang, R. 1997b. Introduction. In Chang (ed) 1997a.
- Cherniak, C. 1981. Minimal rationality. *Mind*, 90(358).
- Cherniak, C. 1983. Rationality and the structure of human memory. *Synthese*, 57(2).
- Cherniak, C. 1986. *Minimal Rationality*. The MIT Press, Cambridge, MA.
- Cohen, L. J. 1992. *An Essay on Belief and Acceptance*. Clarendon Press, Oxford, UK.
- Davidson, D. 1984a. Radical interpretation. In *Truth and Interpretation*. Oxford University Press, Oxford, UK.
- Davidson, D. 1984b. Truth and meaning. In *Truth and Interpretation*. Oxford University Press, Oxford, UK.
- Davidson, D. 2001. A coherence theory of truth and knowledge. In *Subjective, Intersubjective, Objective*. Oxford University Press, Oxford, UK.
- Davidson, D. 2004a. Deception and division. In *Problems of Rationality*. Oxford University Press, Oxford, UK.
- Davidson, D. 2004b. Incoherence and irrationality. In *Problems of Rationality*. Oxford University Press, Oxford, UK.

- Davidson, D. 2004c. Paradoxes of irrationality. In *Problems of Rationality*. Oxford University Press, Oxford, UK.
- Davidson, D. 2004d. Who is fooled? In *Problems of Rationality*. Oxford University Press, Oxford, UK.
- Dennett, D. C. 1987. *The Intentional Stance*. The MIT Press, Cambridge, MA.
- Dougherty, T., editor 2011. *Evidentialism and its Discontents*. Oxford University Press, Oxford, UK.
- Easwaran, K. and Fitelson, B. 2015. Accuracy, coherence and evidence. *Oxford Studies in Epistemology*, 5.
- Egan, A. 2008. Seeing and believing: Perception, belief formation and the divided mind. *Philosophical Studies*, 140.
- Egan, A. 2021. Fragmented models of belief. In Borgoni et al. 2021.
- Elga, A. and Rayo, A. 2021. Fragmentation and information access. In Borgoni et al. 2021.
- Elgin, C. Z. 2017. *True Enough*. The MIT Press, Cambridge, MA.
- Fodor, J. and Lepore, E. 1992. *Holism: A Shopper's guide*. Blackwell, Cambridge, MA.
- Fogal, D. 2019. Rational requirements and the primacy of pressure. *Mind*.
- Foley, R. 1993. *Working Without a Net: A Study of Egocentric Epistemology*. Oxford University Press, Oxford, UK.
- Gabbay, D. M., Johnson, R. H., Ohlbach, H. J., and Woods, J., editors 2002. *Handbook of the Logic of Argument and Inference: The Turn towards the Practical*, volume I of *Studies in Logic and Practical Reasoning*. Elsevier Science B. V., Amsterdam, The Netherlands.
- Gendler, T. S. 2008. Alief and belief. *The Journal of Philosophy*, 105.
- Gigerenzer, G., Todd, P. M., and the ABC Research Group 1999. *Simple Heuristics that Make us Smart*. Oxford University Press, Oxford, UK.
- Greco, D. 2014. A puzzle about epistemic akrasia. *Philosophical Studies*, 167(2).
- Greco, D. 2015. Iteration and fragmentation. *Philosophy and Phenomenological Research*, XCI(3).

- Grim, P. 2004. What is a contradiction? In Priest et al. 2004.
- Hájek, A. 2008. Argument for - or against - probabilism? *British Journal of Philosophy of Science*.
- Harman, G. 2002. Internal critique: A logic is not a theory of reasoning and a theory of reasoning is not a logic. In Gabbay, et al. 2002.
- Harman, G. 2004. The practical aspects of theoretical reasoning. In Mele and Rawling 2004.
- Joyce, J. M. 1998. A nonpragmatic vindication of probabilism. *Philosophy of Science*, 65(4).
- Kindermann, D. and Onofri, A. 2021. The fragmented mind: An introduction. In Borgoni et al. 2021.
- Kolodny, N. 2007a. How does coherence matter? *Proceedings of the Aristotelian Society*.
- Kolodny, N. 2007b. State or process requirements? *Mind*, 116(3).
- Lewis, D. 1982. Logic for equivocators. *Noûs*, 16(3).
- Lord, E. 2014. The coherent and the rational. *Analytic Philosophy*, 55(2).
- MacFarlane, J. 2004. In what sense (if any) is logic normative for thought? Presented at the central division APA meeting.
- Makinson, D. C. 1965. The paradox of the preface. *Analysis*, 25(6).
- Mares, E. D. 2004. *Relevant Logic*. Cambridge University Press, Cambridge, UK.
- Marušić, B. 2015. *Evidence and Agency: Norms of Belief for Promising and Resolving*. Oxford University Press, Oxford, UK.
- Matthewson, J. and Weisberg, M. 2009. The structure of tradeoffs in model building. *Synthese*, 170.
- Meheus, J., editor 2002. *Inconsistency in Science*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
- Mele, A. R. and Rawling, P., editors 2004. *The Oxford Handbook of Rationality*. Oxford University Press, Oxford, UK.
- Mercier, H. and Sperber, D. 2017. *The Enigma of Reason*. Harvard University Press, Cambridge, MA.

- Moran, R. 1997. Self-knowledge: Discovery, resolution and undoing. *European Journal of Philosophy*, 5(2).
- Moran, R. 2001. *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton University Press, Princeton, NJ.
- Murdoch, I. 1971. *The Sovereignty of Good*. Schocken Books, New York, NY.
- Nagel, T. 1986. *The View from Nowhere*. Oxford University Press, Oxford, UK.
- Nickles, T. 2002. From copernicus to ptolemy: Inconsistency and method. In Meheus 2002.
- Norby, A. 2014. Against fragmentation. *Thought*.
- Priest, G. 1979. The logic of paradox. *Journal of Philosophical Logic*, 8(1).
- Priest, G. 1986. Contradiction, belief and rationality. *Proceedings of the Aristotelian Society*, 86.
- Priest, G. 2006. *In Contradiction: A Study of the Transconsistent*. Oxford University Press, Oxford, UK, 2nd edition.
- Priest, G., Beall, J. C., and Armour-Garb, B., editors 2004. *The Law of Non-contradiction: New Philosophical Essays*. Oxford University Press, Oxford, UK.
- Pryor, J. 2018. The merits of incoherence. *Analytic Philosophy*, 59(1).
- Putnam, H. 1963. Brains and behavior. In Butler, R. J., editor, *Analytical Philosophy: Second Series*. Blackwell.
- Raz, J. 2005. The myth of instrumental rationality. *Journal of Ethics and Social Philosophy*, 1(1).
- Rescher, N. 1988. *Rationality: A Philosophical Inquiry into the Nature and the Rationale of Reason*. Clarendon Library of Logic and Philosophy. Clarendon Press, Oxford, UK.
- Rescher, N. and Brandt, R. 1979. *The Logic of Inconsistency*. Blackwell.
- Rips, L. J. 1994. *The Psychology of Proof: Deductive Reasoning in Human Thinking*. The MIT Press, Cambridge, MA.
- Rysiew, P. 2011. Making it evident: Evidence and evidentness, justification and belief. In Dougherty 2011.

- Schotch, P. K. and Jennings, R. E. 1980. Inference and necessity. *Journal of Philosophical Logic*, 9.
- Sellars, W. S. 1962. Philosophy and the scientific image of man. In Colodny, R., editor, *Science, Perception, and Reality*. Humanities Press/Ridgeview.
- Stalnaker, R. 1984. *Inquiry*. MIT University Press, Cambridge, MA.
- Stalnaker, R. 1991. The problem of logical omniscience, i. *Synthese*, 89(3).
- Steinberger, F. 2019. Three ways in which logic might be normative. *Journal of Philosophy*, 116(1).
- Velleman, J. D. 2015. Time for action. In *Possibility of Practical reason, 2nd edition*. Maize Books, Anne Arbor, MI.
- Worsnip, A. 2015. The conflict of evidence and coherence. *Philosophy and Phenomenological Research*, XCVI(1).
- Yalcin, S. 2016. Belief as question-sensitive. *Philosophy and Phenomenological Research*.