

DELIBERATIONS ON DELIBERATION

By

DAVID ANTON JOHNSON

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Philosophy

Written under the direction of

Andrew Egan

And approved by

New Brunswick, New Jersey

October 2020

ABSTRACT OF THE DISSERTATION

Deliberations on Deliberation

by DAVID ANTON JOHNSON

Dissertation Director:

Andrew Egan

Our capacity for reflective agency, the capacity to critically reflect on the contents of our minds and form deliberative judgments on the basis of such reflection, has enjoyed widespread appeal in philosophy. Surprisingly, philosophy has long been without a concrete account of the mental phenomenon by which we endeavor to form reflective judgments. In an effort to fill the deficit, and seeking to offer an adequate account of this phenomenon, which I call *deliberation*, I introduce two explanatory constraints for an adequate account of deliberation. First, an account of deliberation should be sufficiently general so as to provide a unified explanation of various subspecies of deliberation (*e.g.*, theoretical, practical, and moral deliberation). Second, any account should treat deliberation as a mental process in which intentional and reflective agency contribute to the causal structure of the process.

To satisfy these constraints, I offer an account of deliberation that I call the Question-Directed Attitude View (QDAV). According to the view, deliberation is a mental process that has the function of resolving a deliberative state: a unique kind of interrogative attitude consisting of a question and an intentional, attitudinal relation to that question. Deliberative states are resolved relative to the information to which a deliberator reflectively attends. In order to update her reflective information and resolve her deliberative question, a deliberator will shift her attention to different fragments of information. Thus, a crucial component of the view expands on the idea of belief fragmentation by envisioning a deliberator's system of beliefs, desires, intentions, and normative judgments as collections of compartmentalized fragments which might contain distinct, perhaps even conflicting, information.

Drawing on research on iterated belief revision, a formal implementation of QDAV is also developed. The resulting framework models the content of a deliberative state as a partition on a space of possibilities and explores question-resolution using updates to a deliberator's fragmented information structure. The framework helps investigate a crucial distinction drawn by QDAV. Namely, two distinct operations jointly constitute a given update to a deliberator's fragmented information structure: i) the selection of some accessible fragment of information of which to reflectively attend and ii) the integration of that fragment into the deliberator's reflective information. I go on to suggest that norms for deliberation govern the selection operation. I also explore a method for evaluating the way in which a deliberator's information is structured so as to render her an efficient or inefficient deliberator.

The success of QDAV requires that we admit interrogative attitudes into our ontology of the mind. In some sense, this is a radical departure from canonical assumptions within the philosophy of mind, as it calls for the recognition of attitudes bearing non-propositional semantic content. To further motivate recognition of questions as a kind of mental content, I contend that aside from helping explain deliberation, there are other considerations for viewing interrogative attitudes as a useful addition to our theoretical toolbox. To this end, I advance a view of belief-forming processes on which such processes have the function of resolving an interrogative attitude. The view provides for a method of identifying the causes of a belief which are constitutive of belief-formation. I argue that this allows for a useful distinction – one which affords the causal theorist of epistemic basing with a stronger solution to the problem of causal deviance.

Acknowledgements

I owe much gratitude to the community that supported and enriched this dissertation. Of special mention are the teachers and mentors who served as a source of inspiration throughout my philosophical education. The first of these were members of the Amherst College Philosophy Department, where my interest in pursuing philosophy at the graduate level first emerged, and whose support helped bring about this reality. Secondly, I want to acknowledge the Rutgers Philosophy Department and its extended community for providing such a rich intellectual space, as well as those members who live, socialize, and philosophize in New York, for providing many occasions on which to develop these project's ideas.

In particular, I extend a warm thanks to all the members of my committee for their sustained and generous guidance, both intellectual and otherwise. Thank you to Brendan Balcerak-Jackson for agreeing to serve as an external reader, and whose thoughtful feedback on later versions of the project led me to better articulate and locate the project within wider conversations on inquiry and reasoning. I owe much of my philosophical debts to Liz Camp, whose intellectual perspectives animate much of this project, and whose mentorship and support were key to my morale during the writing process. I am grateful to have Susanna Schellenberg as a reader; as I worked through the seeds of this project in my first years as a graduate student, I was especially taken by the idea that a fruitful approach to explaining conscious phenomena might appeal to the mental activity underwriting the respective phenomenon. In many respects, my view seeks to account for the intentional and reflective elements of deliberation in an analogous way. Finally, a special thanks is owed to my advisor, Andy Egan. I am grateful to have learned from his expertise and approach to philosophy, which has trained me to better draw on insights from the philosophy of language in order to enrich my philosophy of mind. His insights at multiple stages of the project offered pivotal direction, and conversation always proved fruitful and enjoyable. To you all, I am extremely grateful to have been extended an invitation to develop my thinking alongside such talented, thoughtful, and kind philosophers.

The development of this project was also undoubtedly influenced by my time spent as a graduate student mentor for the Rutgers Summer Institute for Diversity in Philosophy. I am humbled and grateful for the opportunity to have spent many summers at the “inn” – philosophizing and socializing with so many talented students, undergraduate and graduate students, as well as those faculty members whose accomplishments continue to serve as a source of inspiration. Of course, these experiences would not have been possible without the vision and dedication of its founder, Howard McGary, whose mentorship and friendship I deeply value. His scholarship and activism I consider unmatched.

I must also thank Mercedes Diaz, who has served as a tireless advocate and empathetic friend since the very beginning of my time at Rutgers. Your dedication to the Rutgers Philosophy Department, and to my success as both a philosopher and person, continues to be wholeheartedly appreciated.

Finally, a word is owed to my family and loved ones. I cannot say enough about the patience, friendship, and love my family here in New York has always provided, and the extent to which it has sustained me. Special mention is owed to my mother, Cecilia Jaramillo, who not only imbued in me a profound respect for education from an early age, but also provided me with everything necessary in order to pursue it. You have always been my greatest teacher. Since words will never suffice to express my gratitude, my hope and promise is that I will make you proud. And to Claire, my best friend, in philosophy, and in life. Thank you for your vibrancy, your wit, and your generosity, all of which have carried me through this process.

Dedication

To Tyler
(1985-2014)

Table of Contents

Abstract	ii
Acknowledgements	iv
Dedication	vi
List of Figures	ix
1. Scope and Overview	1
1.1. Chapter outline	1
1.2. Functionalist theory of mind	4
1.2.1. An interpretationist alternative?	7
1.2.2. Broad content	9
1.3. From mental content to deliberative information	10
1.4. Deliberation as a kind of inquiry, reasoning as a kind of deliberation	13
2. The Question-Directed Attitude View of Deliberation	16
2.1. Introduction	16
2.2. Characterizing the phenomenon	17
2.3. The question-directed attitude view of deliberation	23
2.3.1. Questions as content	24
2.3.2. An analysis of epistemic deliberation	27
2.3.3. The general framework	33
2.4. Closing remarks	37
2.4.1. Inferential & non-inferential deliberation	37
2.4.2. Higher-order views of deliberation	44

3. A Fragmented Information Revision Framework for QDAV	48
3.1. Introduction	48
Part I	49
3.2. The naive implementation of QDAV	49
3.2.1. Problematizing the naive implementation	52
3.3. The fragmented implementation of QDAV	54
3.3.1. Fragmented implementation: second pass	62
Part II	67
3.4. Selection operations	67
3.5. Towards a virtue theoretic framework	73
3.5.1. Deliberative efficiency & correctness	75
3.5.2. Deliberative efficiency & length	81
3.5.3. Deliberative efficiency as volume	83
4. A Question-Directed Attitude View for Epistemic Basing	87
4.1. Causal deviance & the epistemic basing relation	87
4.2. Belief-processes as question resolving	91
4.2.1. Questions as mental content	94
4.2.2. Function of belief formation	97
4.3. Question-directedness & basing	102
4.3.1. Epistemic Performances of Belief-Formation	103
4.3.2. Epistemic Performance of Belief-Sustainment	105
4.4. Conclusion	106
References	108

List of Figures

3.1. d during $t_0 - t_4$	51
3.2. Deliberation d of Detective 1	53
3.3. Deliberation d' of Detective 2	53
3.4. a time slice representing information fragments	55
3.5. $i_0 \bullet i_0 = i_1$ (failure to access a new information fragment)	56
3.6. $i_1 \bullet f^1 = i_2$ (updating with f^1 during $t_1 - t_2$)	57
3.7. Deliberative Information Structure STR	60
3.8. deliberative process d^1	60
3.9. deliberative process d^2	60
3.10. What should $i_1 \bullet f^2$ look like?	62
3.11. i_1 & \mathcal{S}_1	63
3.12. updating i_1 with f^2 looks at the smallest sphere in \mathcal{S}_1 consistent with f^2	64
3.13. Which fragment to select first?	68
3.14.	68
3.15. Closeness 1 dictates selecting f^3	69
3.16. d during $t_0 - t_1$, obeys Closeness 1	69
3.17. d during $t_1 - t_2$, obeys Closeness 1	71
3.18. d' during $t_0 - t_2$, violates Closeness 1	71
3.19. $Q^1 \cap Q^2 = Q^{priv}$	74
3.20. STR^a vs. STR^b	76
3.21. STR^c vs. STR^d	77
3.22. STR^a vs. STR^b	79
3.23. Comparing Closeness Scoring 1 - 3	79
3.24. STR^f vs. STR^g	81

3.25. STRC^a vs. STRC^f	83
3.26. STR^g vs. STR^h	85

Chapter 1

Scope and Overview

1.1. Chapter outline

This manuscript is primarily devoted towards sketching out a view of deliberation: those intentional and reflective movements of the mind, or “changes in view”. The championed view is the Question-Directed Attitude View (QDAV), according to which deliberation is a mental process that has the function of supplying intentional and reflective answers to questions. More specifically, the view proposes that deliberation is governed by an interrogative attitude: a mental state with a question as its content. I call the particular kind of interrogative attitude governing deliberation a *deliberative state* and envision it as the analog of an intention. Intentions are characterized by an attitudinal relation to a proposition p , demanding that the occupier of the intention do what she can to make p true. Similarly, occupying a deliberative state demands that the deliberator do what she can to resolve whatever question is the content of her deliberative state.

The view is presented in more detail in Chapter 2: The Question-Directed Attitude View of Deliberation. There, I offer a pre-theoretic characterization of deliberation and make the following observations: deliberation occurs across a host of different domains of thought, operating on different kinds of attitudes, and although deliberation involves a sequence of transitions between attitudes, deliberation and non-deliberation can share the same sequence of transitions between attitudes. These observations motivate two adequacy constraints for an account of deliberation. We require a unified account of deliberation that is deployable across different domains of thought (*Neutrality Constraint*). Additionally, our account should analyze deliberation as a mental process in which certain intentional and reflective elements contribute to the causal structure of the process (*Intentional Reflective Process (IRP) Constraint*).

QDAV satisfies both the Neutrality and IRP Constraints. Its success rests on three core features of the view: i) the use of an intentional interrogative attitude (a deliberative state) to specify deliberation's functional role, ii) the set-theoretical modeling of the contents of deliberative states and the contents of propositional attitudes, and iii) the fragmentation of a deliberator's system of propositional attitudes. QDAV addresses the Neutrality Constraint by envisioning all deliberative processes as sharing a single functional relation that holds between the deliberator's propositional attitudes and her deliberative state. (i) and (ii) allow us to represent this relation as one which holds between the semantic contents of a deliberator's propositional attitudes and the semantic content of her deliberative state. The relation causes, or tends to cause, her propositional contents to change in such a way as to resolve the content of her deliberative state. Of course, the devil is in the details. For now, we need only note that QDAV's analysis of deliberation doesn't discriminate between different kinds of deliberation.

With respect to the IRP Constraint, QDAV posits that deliberative states are resolved relative to those propositional attitudes that enjoy the deliberator's conscious, or reflective, attention. In light of (i), deliberative states are intentional, meaning that a deliberator will try to bring about the resolution of her deliberative state. Hence, she will intend to update the contents of those attitudes which enjoy her conscious attention (her "conscious" information) in such a way as to resolve her deliberative state. By way of (iii), QDAV treats a deliberator's system of propositional attitudes as a collection of compartmentalized fragments that might contain distinct, perhaps even conflicting, information. This serves to draw a distinction between the fragment to which she consciously attends and those fragments to which she has not yet attended but are accessible to her. So long as a deliberator can exercise intentional control over which fragments of information she accesses in order to try and resolve her deliberative state, QDAV respects the causal roles that intentional and reflective agency plays in the deliberative process.

Chapter 3: A Fragmented Information Revision Framework for QDAV provides a formal implementation of QDAV by drawing on work in the belief revision literature. The implementation is not offered as the only formal framework for QDAV (or even the best). Instead, the purpose is twofold: make good on the promise that QDAV satisfies the Neutrality Constraint, and illustrate the kind of novel normative issues that arise on a proper understanding of deliberation. As just discussed, QDAV purports to satisfy the Neutrality Constraint by relying on

a set theoretic modeling of content that abstracts away from the details particular to any one kind of deliberation. To help substantiate the claim, it might help to have a formal framework that cashes out the details of a set theoretic modeling of the states and relations involved in the process of deliberation. The framework also serves an important purpose in highlighting a crucial distinction drawn by QDAV. Namely, two distinct operations jointly constitute an update to a deliberator's fragmented information structure: i) the selection of some accessible fragment of information of which to consciously attend and ii) the integration of that fragment into the deliberator's conscious information. The distinction helps us investigate the normative dimension of deliberation. I suggest that norms for deliberation govern the selection operation. I also explore a method for evaluating the way in which a deliberator's information is structured so as to render her an efficient or inefficient deliberator.

The success of QDAV requires that we admit interrogative attitudes into our ontology of the mind. In some sense, this is a radical departure from traditional assumptions within the philosophy of mind, as it calls for the recognition of attitudes bearing non-propositional semantic content. To further motivate recognition of questions as a kind of mental content, I contend that aside from helping explain deliberation, there are additional reasons for viewing interrogative attitudes as a useful addition to our theoretical toolbox. To this end, Chapter 3: A Question-Directed Attitude View for Epistemic Basing advances a view of belief-forming processes according to which such processes have the function of resolving the question that is the content of an interrogative attitude. The view provides for a method of identifying the causes of a belief which are constitutive of belief formation. I argue that this allows for a useful distinction – one which affords the causal theorist of epistemic basing a more satisfactory response to the problem of causal deviance.

To better situate the above issues, a few preliminary remarks are in order. The remainder of this introductory chapter attends to three different issues with the intention of clarifying terms and drawing out the core assumptions that have shaped my deliberations on deliberation.

1.2. Functionalist theory of mind

I assume a functionalist view of the mind according to which types of mental states can be individuated by the functional role the state occupies in a subject's mental economy. I'll also assume that types of mental processes –causally structured collections of mental states– can be functionally individuated. Hence, I take it that a fruitful approach will treat deliberation as a mental process with a specified functional role. Moreover, the functional role should properly explain the causal structure of deliberation and should generally cohere with our armchair observations and analysis of possible cases of deliberation. The kind of view I am after is a view of deliberation *tout court*. It won't just account for the various kinds of deliberation we engage in on a day-to-day basis (*e.g.*, theoretical, practical, and moral deliberation). If one day we discover intelligent life on mars, then the account I am after should be able to account for the deliberation of the Martians as well. The functionalist leanings driving my approach are very much in the spirit of the analytic functionalist who privileges folk psychological explanations and conceptual analysis of possible cases to identify the functional role of a state (see Shoemaker, 1984; Lewis, 1972; Jackson, 1996). However, for reasons soon to be discussed, some ways of implementing analytic functionalism might not fully map on to my approach.

With respect to contentful states, I'll largely assume that such content is representational and contributes to the internal, causal relations which the state stands in. Of course, two token mental states might be type-identical while differing in content. However, once we recognize that content makes a causal contribution, we open up the possibility that mental states with markedly different kinds of representational content will not share the same functional role, and hence will not be type-identical. I think we will find this to be the case when we look at the difference between propositions and questions. Crudely, a proposition is, or picks out, some possible state of affairs. This seems fundamentally distinct from a question which merely draws a distinction between possible states of affairs. The question *will it rain?* distinguishes between the state of affairs in which it will rain and the one in which it won't.

Given this difference between propositions and questions, we should not be too surprised if it turns out that mental states with propositional content must have different functional roles than mental states with questions as content. Of course this will turn on whether we decide

to recognize interrogative attitudes in the first place. And it is here where my relationship to analytic functionalism becomes complicated. Analytic functionalists have, by and large, been more concerned with functionally individuating mental-state-types rather than process-types. My interests are comparatively more process centered. Moreover, analytic functionalists tend to tacitly assume that all semantic content is propositional. Their general strategy has been to use conceptual analysis to identify a functional role that best coheres with folk psychology. Crucially, folk psychological platitudes and explanations tend to be cast in terms of behavioral and motor responses, perceptual inputs, mental causation, and *propositional attitude attributions*. Where does this leave deliberative states and other interrogative attitudes? Are there folk psychological platitudes concerning them? It's unclear.

Here are some potential platitudes to consider: *deliberating, reasoning, considering, or wondering whether p tend to cause one to believe or disbelieve p*. Do the platitudes provide direct support for an interrogative attitude corresponding to any of the processes of deliberating, reasoning, considering, or wondering? In other words, do any of the platitudes presuppose an attitudinal relation to questions as semantic content? Or do the platitudes only provide direct support for the propositional attitude of belief? It might be that the platitudes presuppose propositional attitudes in order to convey a causal relation between a mental process and belief without any commitment to the existence of an interrogative attitude.

If there are folk psychological platitudes about interrogative attitudes, then the analytic functionalist can view my project as an endeavor to use those platitudes to reveal the functional role of deliberative states. This provides a fairly straightforward understanding of how I use various cases. See for instance, the case observations I draw in Section 2.2 of Chapter 2 to motivate the Neutrality and IRP Constraints. We might think of those armchair observations as making salient certain platitudes about the kind of interrogative attitude that I've been referring to as a "deliberative state". In which case, the next chapter can be read as appealing to these platitudes to identify the correct functional role of deliberative states. Suppose, however, there are no platitudes which presuppose an interrogative attitude. Then, it's likely that an analytic functionalist will only recognize the class of interrogative attitudes if doing so can help her make better sense of the folk psychological platitudes she does recognize. The analytic functionalist can then view my project primarily as a humble endeavor to establish the interrogative

attitudes as a genuine type of mental state.

That interpretation of the project will lean heavily on evaluative platitudes about propositional attitudes. Contemplate platitudes like *beliefs formed on the basis of deliberate reflection are rationally evaluable in a way that beliefs formed on the basis of unconscious processes are not*, and *agents are more responsible for beliefs formed on the basis of deliberate reflection*.¹ Assuming she is not already presupposing the existence of interrogative attitudes, the analytic functionalist won't treat these platitudes as direct evidence for how to fix the functional role of a deliberative state. But, she can be expected to treat these platitudes as direct evidence for functional relations that hold between the process of deliberation and the attitude of belief. If it turns out that the best explanation of those functional relations posits a certain attitudinal relation to a question (*i.e.*, a deliberative state), then that will be some reason to admit the interrogative attitudes into our ontology of the mind.

All this is just to say that for the analytic functionalist who is skeptical of treating questions as a kind of mental content, it will be critical to consider Chapters 2 and 4 in tandem. Both those chapters can be read as positing particular kinds of interrogative attitudes –so-called “deliberative states” and “question states”– in order to provide explanations that better cohere with a version of folk psychology which does not presuppose interrogative attitudes. Perhaps any remaining skepticism about interrogative attitudes will be dispelled, when considering other projects that demonstrate the usefulness of interrogative attitudes (See Friedman, 2013, 2019). When that occurs, Chapter 2 will stand on more tenable grounds.²

While I have been focusing on the status of the interrogative attitudes, much of what I have

¹One might take issue with the idea that there are folk psychological platitudes which involve normative-predicate attribution. I think folk psychology helps itself to many normative notions, like rationality or the state of knowing. Hence, it's not unreasonable to think that there might be some folk psychological platitudes about normativity. Nonetheless, if we must restrict ourselves to non-normative platitudes, then I should be able to secure my point by modifying the two platitudes above as follows: *beliefs formed on the basis of deliberate reflection are judged to be rationally evaluable in a way that beliefs formed on the basis of unconscious processes are not*, and *agents endorse beliefs formed on the basis of deliberate reflection*. Neither of those platitudes concern anything normative, they only concern an agent's attitudes towards normative notions.

²It is worth noting that interrogative attitudes might not be the only kind of non-propositional content which contribute to a more complete theory of mind. For the kind of non-propositional, dispositional states I have in mind, see (Camp, 2009, 2013, 2020). While characterizations and perspectives prove illuminating, I wonder to what extent certain kinds of higher-order interrogative attitudes might secure some of the same explanatory aims. In any event, a promising way to supplement my main argumentative move might appeal to evidence of other kinds of non-propositional mental representation, argue for their explanatory power *viz-a-viz* folk-psychological phenomena, and then, finally, treat the resulting conclusions as evidence for a more promiscuous approach to the kinds of mental representations we recognize.

said also holds true for the structure of a deliberator's propositional attitudes. The analytic functionalist will be faced with the choice of whether to adopt a version of folk psychology which presupposes fragmentation (this strikes me as most plausible) or a version which does not. The chosen option will determine whether we can help ourselves to fragmented attitude structures, or whether we need to posit them as theoretical constructs for explanatory purchase.

1.2.1 An interpretationist alternative?

Of course, functionalism is not the only basis for the general argumentative move I want to make to motivate QDAV. Consider an interpretationist position according to which facts about the contents of one's mind are determined by the agent's complete behavioral history along with facts about how to construct the best folk-psychological explanation about that behavioral history. Traditionally, interpretationism has been most concerned with determining the contents of an agent's beliefs and desires by using belief and desire attributions to explain one's behavioral history (Lewis, 1983, 1974; Dennett, 1988; Davidson, 1973). However, we might expand the interpretationist program to include deliberative states and fragmented belief and desire systems, wherein, one will qualify as having a certain deliberative state and a certain fragmented belief and desire system just in case such attributions are part of the best folk-psychological explanation of her behavioral history.

However, on this picture, all it would mean for an agent to have a certain deliberative state and fragmented system of attitudes is for her behavior to be interpretable in a certain way. Part of the interpretationist strategy involves explaining an agent's behavioral history by way of a charitable interpretation of her beliefs and desires. For reasons Lewis (1974) points out, a charitable interpretation will explain the agent's behavioral history by attributing to her those beliefs and desires which it would be most reasonable for her to have (as opposed to attributing to her true beliefs). By adding deliberative states and fragmentation to the interpretationist program, we provide additional mechanisms for securing a charitable interpretation. My worry is that those mechanisms go too far. Surprisingly, the more charitable interpretations provided by deliberative states and fragmentation cause trouble for interpretationism by precluding the possibility of error and increasing content indeterminacy. Below, I will briefly run through how accommodating for fragmentation raises these two concerns. (Analogous conclusions can

be drawn about deliberative states.)

Suppose Claire usually, but not always, wears her seat-belt while driving. Without fragmentation, a charitable interpretation that coheres with Claire's behavior is likely to interpret her as desiring to-wear-her-seat-belt-when-driving. A charitable interpretation is also likely to interpret her as having the necessary beliefs to satisfy her desire to-wear-her-seat-belt-when-driving. After all, she usually succeeds in satisfying her desire, indicating that she has all the relevant beliefs about seat-belts. Moreover, beliefs about how to wear a seat-belt certainly seem like they are among the set of beliefs that would be most reasonable for one to believe. Notice, though, that because Claire sometimes does not wear-her-seat-belt-when-driving, the interpretationist will be forced to recognize instances in which Claire fails to behave in accordance with the charitable interpretation of her beliefs and desires. But, perhaps some level of error is to be expected. Actual agents –the best of them– are prone to cognitive malfunction.

Now suppose the interpretationist can use fragmented desire attributions to interpret Claire's behavior. The most charitable interpretation will likely interpret her as having the same beliefs as before. Using fragmented desire attributions, we can interpret her as having a desire to-wear-her-seat-belt-when-driving which is contained in a fragment that is only active on those occasions in which she wears her seat-belt when driving. The most charitable interpretation of Claire should interpret her in just this way as it renders her seat-belt-wearing behavior error-proof. Using this method, it might be possible to interpret her entire behavioral history using fragmented desire attributions compartmentalized in such a way that she always acts in accordance with her beliefs and (active) desires. But, interpretations that shield Claire's behavior from the possibility of error are much too charitable.

Another cause for concern is the multitude of different ways Claire's beliefs and desires might be fragmented. Suppose that every time she drives on a Sunday she wears her seat-belt and every time she drives to the airport she also wears it. Should we interpret her as having a desire to-wear-her-seat-belt-when-driving which is contained in a fragment that is activate when driving on Sundays and as having a fragment containing a similar desire which is active when driving to the airport? What about a single fragment that has the desire and which is active when driving on a Sunday and when driving to the airport? Without some constraints that explicitly limit the range of candidate fragmented attitude structures, implementing fragmentation on the

interpretationist program increases the risk of content indeterminacy. In light of this concern, and the concern about the possibility of error, I prefer to treat the explanatory power afforded by deliberative states and a fragmented attitude structure as evidence for a more substantive view about mental content and organization.

1.2.2 Broad content

Adopting the kind of functionalist framework I have in mind might seem to result in an unwelcomed restriction on the kind of questions one might deliberate about. Here, I have in mind arguments that purport to show a tension between functionalism and mental states with so-called *broad content* that involves reference to a subject's environment. The worry is that broad-content mental states seem unsuitable for functional analysis using only causal relations to internal inputs and outputs. Similarly, it has been argued that broad-content mental states can't contribute to functionalist explanations of an agent's behavior (Fodor, 1987). If these kind of worries are on track, then perhaps adopting functionalism will have the consequence of precluding deliberative states with broad content. So for instance, one couldn't occupy a deliberative state about a question like *is the cup filled with water?* (where *water* refers to H₂O). But surely we should recognize that sometimes deliberators engage broad-content questions for which the answers will have broad content. The most straightforward way of capturing that on the QDAV framework would seem to recognize broad-content deliberate states.

In response, we might move to a version of functionalism which allows for the specification of functional roles in terms of relations to both internal and environmental inputs and outputs (Block, 1990). This would allow for the proper individuation of broad-content mental states. In fact, we might even distinguish between mental states *M* that stand in causal relations to environmental inputs and outputs and mental states *M** that stand in causal relations to the *M* states, taking the *M** states to serve as broad-content mental states. On this picture, a broad-content mental state about water will be the state of having a certain belief-role-state which is caused by H₂O. As (Jackson and Pettit, 1988) note, the strategy should let functionalism accommodate content that is essentially broad. However, to simplify our discussions I will proceed as if a mental state's semantic content will require less elaborate treatment.

1.3. From mental content to deliberative information

Generally, I use “mental state” to refer to a contentful mental state. The two kinds I’ll primarily be referencing in this manuscript are deliberative states and propositional attitudes. Deliberative states, which have questions as content, are the states that govern deliberation. Although “interrogative attitude” is the standard way of referring to question-directed mental states, I’ll only use “attitude” to refer to an attitude with propositional content. Here, I am operating with a fairly loose notion according to which propositions need not be the bearers of truth and falsity. Instead, I only require that propositional content obeys compositionality and that its semantic features provide some readily identifiable satisfiability or correctness conditions by which to evaluate propositional attitudes.

Perhaps you are one of those unfortunate philosophers whose restrictive metaphysics leaves no room for properties or facts corresponding to love, justice, personhood, or unicorns. Even so, it would be rather far-fetched to suppose that we can’t successfully deliberate, or otherwise think, about such things. For this reason, we’ll operate with a notion of proposition which can serve as the content of thought. More specifically, a proposition will be the content of an attitude which is evaluable, due in part, to the semantic properties of its content (this is not meant to exclude other forms of evaluation). For example, someone who denies metaethical realism might take the position that moral content fails to reflect “objective” reality. It’s still coherent for this individual to think that someone who judges murder morally impermissible would be making a mistake to also judge the murder of one’s enemies as morally permissible.

We will take the content of a propositional attitude to be representable using some set-theoretic space of possibilities. The simplest construction of possibilities will be cast in terms of possible worlds, but it’s almost certain that we will need to complicate the possible worlds framework to accommodate the myriad of different kinds of propositional content. For instance, attitudes with *de se* content will likely require something like a set of pairs of possible worlds and individuals. I have no strong commitments about how best to model different kinds of propositional content. However, I do want to make a general point about the expressive power of certain modelings. Consider a space of possibilities in which each point is a centered world: a world-individual pair. That construction of logical space will let us model both *interesting*

and *boring de se* contents (Egan, 2018). An interesting *de se* content, like *I am thirsty*, carves out a set of possibilities that cuts across world-mates.³ A boring *de se* content, like *Louis Armstrong's pants are on fire*, does not cut across world-mates; it could just as easily have been modeled by a construction of logical space on which individual centers are entirely absent. Moreover, as Lewis (1979) notes, any content that can be adequately modeled using possible worlds alone can be adequately modeled by centered worlds as boring *de se* content.

The observation that boring *de se* content can always be used to model standard possible-world content holds because we can use any world w to construct a centered world that contains w . This is easily checked: for any world w , all we need is a non-empty set of individuals to construct centered worlds by pairing w with those individuals. So, in essence, a centered world contains everything that a possible world does plus something extra (a centered individual). The point generalizes: if we take a space of possibilities, we can always take its most basic set-theoretic elements and add additional elements to redefine points in logical space. The new construction of logical space can only increase in expressive power, as it will still be able to model all the same content as the original. This point is important when it comes to thinking about the informational contents of what I call a *deliberative domain*.

A deliberative domain will be populated by a cluster of propositional-attitude-types. It might be that we initially modeled the content of each attitude-type in a domain using different spaces of possibilities. Nonetheless, we know now that we can always take the most basic set-theoretic entities of each respective possibility space, and construct a new, richer space capable of modeling the joint contents of all the different attitude-types of the domain. That joint content will comprise the information of a deliberative domain. Think of the totality of a domain's deliberative information as all the possible information which might ever be relevant to whatever deliberation occurs in the domain. Here's another way of arriving at the same point. Whatever deliberation occurs in a domain will consist only of a deliberative state and the domain's attitude-types. Hence, because deliberative information captures the joint contents of the attitude-types populating a domain, a domain's deliberative information can represent the

³The proposition *I am thirsty* will consist of a region of logical space X for which one of the points will consist of the actual world and the author. But assuming there is someone in the actual world who is not thirsty, there will be a point consisting of the actual world and an individual which is not contained in X .

various semantic contents which occur during deliberation.

For instance, suppose the domain of practical deliberation involves beliefs, desires, and intentions. Suppose also that we start by modeling belief-content using centered worlds (world-individual pairs). We then decide to model desire-content using possible worlds, and intention-content using world-hyperplan-pairs (see 2.3.3 for discussion of hyperplans). What will the content of the practical domain look like? To answer that question, we start by looking at the basic set-theoretical entities used to model the three kinds of content in the practical domain: worlds, individuals, and hyperplans. Now all we need to do is construct a new space of possibilities where each point will be a world-individual-hyperplan triple. That space of possibilities defines the deliberative information of the practical domain, and it will be sufficient to model any combination of beliefs, desires, and intentions.

One might be left with the impression that I am individuating deliberative domains by the cluster of attitude-types which populate the domain. However, that's not the whole story. The task of the preceding paragraphs was only to introduce a domain's deliberative information as the space of possibilities which models the joint content of the domain's cluster of attitudes. So, while it is true that deliberative information is determined solely by the attitude-types that populate a domain, that won't be the case for deliberative domains. Deliberative domains are individuated by the attitude-types which populate the domain *and* causal relations that determine which of its attitude-types can serve as inputs to the domain's deliberative process and which can serve as outputs.

As an example, recall that on our toy account of the practical domain, the domain is populated by beliefs, desires, and intentions. This alone won't be enough to individuate that domain. Imagine we ascribe to some strong form of doxastic voluntarism, according to which it is possible to deliberate about what to believe on the basis of one's beliefs, desires, and intentions. Both the practical domain and a voluntaristic-epistemic domain are populated by the same three attitude-types. Does this force us to view the two domains as identical? Not necessarily. After all, we might think that practical deliberation terminates in a intention, while voluntaristic-epistemic deliberation terminates in a belief. Thus, we will want to distinguish deliberative domains on the basis of the attitude-types which populate them and the causal-relations between those attitude-types during domain-specific deliberation.

The last issue I'll raise about deliberative domains and information concerns the modeling of the deliberative states that I have tasked with governing the deliberative process. Deliberative states have questions as their semantic content. Throughout the manuscript, I model questions as partitioning the construction of logical space on which the question is defined. Hence, to determine the content of a deliberative state, one needs to know the deliberative domain it inhabits. Crucially, QDAV will fix the functional role of a deliberative state and the deliberative process by relying on semantic relations between a question in a deliberative domain and the information of that domain. The semantic relation between deliberative questions and what I am calling "deliberative information" might help ease concerns that I am abusing the term "information" by using it to include the contents of pro-attitudes.

With respect to pro-attitudes, like desires and intentions, the so-called direction of fit is from mind-to-world. So, for those whose default notion of information is one that represents mind-independent features of the world, it might seem inappropriate to talk about pro-attitude content as "information". My usage of "information" is only meant to pick-out those semantic contents which are relevant for answering certain questions. Because we sometimes deliberate about questions about what to do, or even how to feel, we will sometimes need to think about the contents of pro-attitudes as information in the way I think about information –as that which can play a role in resolving a question.

1.4. Deliberation as a kind of inquiry, reasoning as a kind of deliberation

Deliberation, as I use the term, refers to a class of thought which is part of the broader category of inquiry. Inquiry, as I understand it, is question-directed thought. Separating deliberation from non-deliberative inquiry is a tricky task. In Chapter 2, I offer two marks of the deliberative. Deliberation is intentional and reflective in ways that most other forms of inquiry are not. Specifically, a deliberator not only intends to resolve a question but intends to do so by consciously attending to the contents of her mental states. We might also consider additional ways of distinguishing deliberation from non-deliberative inquiry. All deliberation is inherently normative, while at least some non-deliberative inquiry is not. Consider, for instance an inquiring

act of wondering about tigers. Quite plausibly, there aren't good or bad ways to wonder about tigers. It might strike us as more peculiar to wonder whether more tigers have an even number of stripes than it is to wonder how much the typical adult tiger weighs in the wild. Nonetheless that doesn't make it any worse of a wondering.

Of course, some non-deliberative inquiry is normative. Various belief-forming processes are normatively evaluable without qualifying as deliberation. At least on the view of belief-forming processes developed in Chapter 4, belief processes are a kind of question-directed thought and will hence qualify as a form of inquiry⁴ Consider for instance a visual belief-forming process. Plausibly, such a process will likely involve a hierarchy of sub-processes which use visual cues. For present purposes, we can take visual cues to be anything from an "implicit assumption", triggered by visual stimuli, to Bayesian probabilities. Using visual cues, these collections of sub-processes will inferentially construct features of one's visual scene from raw perceptual stimuli. But, why do we have the visual cues we do?

According to my Question-Directed View of belief-processes, the function of a belief-forming process will be to resolve a question. Properly identifying the operative question in beliefs formed on the basis of, for example, visual perception will surely prove difficult (it's a task best left to a vision scientist). But it will resemble something like *what is my visual scene like?* or *what visual features does my environment have?*⁵ If our visual system is marked by one such question, then the visual scene constructed by a visual belief-forming process should serve to answer that question. Hence, our visual systems exhibit the visual cues they do in virtue of the function of those processes. As an example, consider the visual cue that tells us that lighting likely comes from above or that smaller objects in the visual field are closer in proximity. I take such cues to be evidence that some of the features which the visual question ranges concern the orientation of local lighting sources.

Visual belief-forming processes are clearly normatively evaluable. But such processes need

⁴I suspect I have a non-standard view of the nature of inquiry which some might find too permissive. For a more standard treatment see (Friedman, 2019)

⁵Specifying the right question which drives our belief-forming processes is challenging because doing so requires that we look at one of the italicized interrogatives and move to its extension. In other words, we need to specify a question that will have the following form. Given visual features f_1, f_2, \dots, f_n , $\langle \text{Does my visual scene have } f_1, f_2, \dots, \text{ or } f_n? \rangle$ The challenge lies in determining the range of candidate visual features which the visual scene of a normal human could have. Clearly, that is a task best left to the vision sciences.

not qualify as deliberation because they are often sub-personal.⁶ An important mark of deliberation is that its normativity is also distinctly personal. When deliberation is done well, the deliberator is praiseworthy. When her visual system performs well, she will be fortunate but not worthy of praise in the same way.

Thinking of deliberation as a form of inquiry which renders the whole agent normatively evaluable is also consistent with the idea that reflective reasoning is a form of deliberation. Reasoning, as I understand it, involves an intentional and reflective movement of the mind, characterized by an inferential sequence between sets of propositional attitudes. This sequence will be causally explained by the reasoner consciously attending to logical or semantic relations between the contents of her attitudes. As it turns out, an important feature of QDAV is that it both respects the privileged place which reflective reasoning has occupied in philosophy while also expanding the kind of phenomenon which hold a similar normative significance.

⁶Though this need not always be the case. So long as we take visual belief-formation to involve a hierarchy of sub-processes, we can define a range of different collections of sub-processes which terminate at different points in the hierarchy. This might allow for question-directed visual-belief formation both at the personal and sub-personal levels. If, on the chance, that a full-blown, personal level visual belief-forming process occurs, I don't necessarily have reservations against typing that process as deliberative or quasi-deliberative.

Chapter 2

The Question-Directed Attitude View of Deliberation

2.1. Introduction

Our capacity for reflective agency –that is, the ability to critically reflect on the contents of our minds in order to form deliberative judgments– has enjoyed widespread appeal in philosophy. Aristotle called upon such faculties in order to construct a notion of moral responsibility; Descartes’ *Meditations* are predicated upon our ability to form deliberative, reflective beliefs; Kant grounds normative demands upon the reflective will. While the traditional philosophical import of reflection is not without detractors, the capacity for reflection continues to play important roles in epistemology, ethics, and the philosophy of action, and the subjects of practical and theoretical reasoning are arguably devoted to uncovering the norms governing acts of reflection.¹ However, the many appeals to our capacity for reflection have been disproportionate to the efforts devoted towards explaining the mental phenomenon associated with our usage of that capacity. What, precisely, is the nature of the phenomenon which underwrites an endeavor to form a reflective judgment?

In what follows, I refer to this phenomenon as *deliberation*. In the next section, I try to pin it down, characterizing it as the intentional and reflective mental process whereby an agent transitions between sets of propositional attitudes. Additionally, I also introduce two constraints which a proper explanation of deliberation should satisfy. In Section 2.3, I examine the ubiquitous yet nebulous notion that deliberation somehow involves questions.

¹See Hilary Kornblith’s *On Reflection* (2012) for a well-argued critique of various attributes philosophers typically attribute to reflection. To briefly gesture at a few examples of the role reflection plays in contemporary discussions, Ernest Sosa in *A Virtue Epistemology* (2007) elevates reflective knowledge above animal knowledge (*i.e.*, non-reflective knowledge) and prizes reflection, in part, for its relation to epistemic agency. Metanormative constructivists, like Christine Korsgaard (2009) and Sharon Street (2008), have proposed that our capacity for reflective agency grounds practical, moral, and perhaps even our epistemic reasons, while David Enoch (2011) argues that the possibility of deliberation supports a robust metanormative realism.

Once properly refined, the notion helps shape a promising account of deliberation, the Question-Directed Attitude View of Deliberation (QDAV) which satisfies the two proposed theoretical constraints. According to the view, deliberation is a process whereby agents bring to bear conscious attitudes to intentionally resolve deliberative states, a distinct kind of non-propositional mental state that has a question as content. In Section 2.4, I discuss some limitations of views of reflective reasoning and views which rely on higher-order propositional attitudes. Doing so helps highlight additional advantages of QDAV by discussing some limitations.

2.2. Characterizing the phenomenon

Deliberation is a movement of the mind whereby an agent adopts, revises, or sustains a view. Although deliberation might not always result in a new view, we might borrow a phrase from (Harman, 1986) and say that deliberation involves a “change in view”.² What is characteristic of deliberation is that the change in view occurs on the basis of the deliberator’s i) intending to undergo a change in view by reflecting on mental contents and ii) proceeding to reflect on those contents³. I will present three examples of deliberation and suggest that some rather non-controversial observations motivate two constraints on a theory of deliberation: a theory should not appeal to any features specific to one domain of thought in which deliberation occurs (Neutrality), and a theory should specify the functional roles of the deliberator’s intentional and

²Harman’s use of “view” is broad so as to include not just (theoretical) views about some descriptive state of affairs but also (practical) views about what to do. Changes in one’s theoretical view will correspond to a change in belief while changes in one’s practical view will correspond to a change in intention. We might further broaden the term to include other kinds of views like moral, aesthetic, and religious views. This will require that we use different clusters of attitudes to underwrite the different kinds of views one might have (see the discussion on deliberative domains in this chapter).

³I will use “reflective” and “conscious” and their derivatives fairly interchangeably. Much more could (and should) be said about what distinguishes a reflective attitude, as well as how this differs and relates to the process of reflection, whereby an attitude becomes reflective. There are different ways to address these issues. I will briefly mention two, the *state* and *resource conceptions*. According to the state conception, in order for some state, *m*, to be reflective, one must occupy a state that is about *m*, a so-called higher-order mental state. From my vantage point, the most plausible version of the state conception will involve a higher-order *de se* state, the content of which references *m* and the subject itself. So reflection will only be possible for creatures that possess the cognitive capacity for self-reference by way of some self-referring concept (Burge, 1996).

The resource conception understands reflection by treating conscious attention as a limited cognitive resource which, when employed by a mental state, contributes to the overall subjective experience of the agent (Peacocke, 1998). I favor this conception, which I adopt in Section 2.3, because it facilitates talk of reflecting on, or consciously attending to, a state as well as a state being reflective or conscious. However, QDAV can be easily retrofitted to adopt a state conception.

reflective agency (Intentional-Reflective Process Constraint). Let's begin by examining how the Neutrality Constraint drops out fairly quickly of the following three cases:

TOURIST: A tourist endeavors to determine her current location relative to a museum she intends to visit. Aware that the museum is in the northern region of the city, she starts scanning the section of the map representing the northern region while trying to recall as many of the various streets and landmarks she recently passed. She identifies one point on the map representing an intersection which she is aware of having recently passed. The map places the intersection just north of the museum. Certain that she had been approaching the museum from the south, she concludes that she must have walked past the museum and forms the belief that she is just a few blocks north of it.

ATTORNEY: An attorney has the option of filing a certain motion for a particularly nuanced case. She is deeply invested in the outcome of the case and spends considerable amount of time reflecting on the possible outcomes. She is fairly certain that if she does not file the motion, it will provide her client with strong grounds to appeal any potential unfavorable judgment. On the other hand, not filing might have negative consequences for her career and reputation. After carefully considering the features of the case, she concludes that there are other plausible grounds by which her client might appeal, and that the likelihood of the client needing to appeal does not warrant the potential risk to her professional reputation. She decides to file the motion.

ETHICS STUDENT: A student sits deep in thought considering whether it is ever permissible for a doctor to take an innocent patient's life in order to save five patients who are in need of organ transplants. She has a strong affective response when reflecting on the case, and treats it as defeasible evidence for a moral judgment. But, she is aware that her prior reflections about a related case might commit

her to judge that it would be permissible for the doctor to sacrifice the one for the many. After further reflection however, she identifies a noteworthy difference between the cases and concludes that while it is sometimes permissible to sacrifice the one for the many, it is impermissible to do so when the death of an innocent is constitutive of the manner in which the many are saved.

All three cases involve a transition between sets of attitudes, or more precisely, a sequence of such transitions, until some final judgment has been reached. The lost tourist experiences a sequence of transitions between doxastic attitudes culminating in a belief about her location relative to the museum. The attorney forms an intention by way of a sequence of attitude-transitions involving various doxastic and pro-attitudes, and the student forms a moral judgment by way of a sequence of attitude-transitions involving doxastic and moral attitudes. While the kinds of attitudes involved in each sequence differ, the attitudes all seem to share propositional content.⁴ In this sense, all three sequences contribute to a “change in view”. Each change in view also exhibits the hallmarks of deliberation insofar as the subject endeavors to experience a change in view by reflecting on the contents of her mind and does so in fact reflect.

We observe, then, that deliberation can occur in at least three domains of thought (the epistemic, practical, and moral), where a domain is individuated by the kinds of propositional attitudes which underwrite the thoughts and views of that domain. This observation should incline us to pursue a unified theory of deliberation which explains an intentional and reflective sequence of transitions between propositional attitudes in a domain-neutral way:

Neutrality Constraint: A theory of deliberation should not appeal to features specific to a domain of thought in which deliberation occurs.

Of course, the force of Neutrality depends on the extent to which we recognize a single mental phenomenon (deliberation) underlying the three cases. What reason do we have to

⁴If we insist on restricting propositions to purely descriptive information about “objective” features of the world, then some metanormative commitments will preclude the possibility of propositional attitudes with normative content. Nonetheless, it is hard to deny that normative judgments possess propositional-like content: contents which can be accepted or rejected, embedded in complex, mixed contents, and which can play a role in inference. I take this to be symptomatic of a shared discursive representational format. My reader is free to interpret my use of “propositional attitude” as an umbrella term.

think this? I will not provide any definitive reasons here, but I will propose that there are some important commonalities between the three cases and that a straightforward way of explaining those commonalities is to classify the intentional and reflective changes in view exhibited in by the three cases as belonging to the same kind of mental process. To draw out some important commonalities across the three cases, contrast each original case with a deviant version fitting the schema below:

Of course, the force of Neutrality depends on the extent to which we recognize a single mental phenomenon –deliberation– underlying the three cases. What reason do we have to think this? I will not provide any definitive reasons here, but I will propose that there are some important commonalities between the three cases and that a straightforward way of explaining those commonalities is to classify the intentional and reflective changes in view exhibited in by the three cases as belonging to the same kind of mental process. To draw out some important commonalities across the three cases, contrast each original case with a deviant version fitting the schema below:

Deviant Case Schema: Similar to the original case, with the following exception. A mad scientist who keeps a basement full of brains-in-vats conspires with an evil demon to have the subject stung by a mosquito carrying a rare disease. Just after the subject endeavors to reflect and form a judgment, she is stung and is immediately stricken with a hallucinatory fever. During this fever, she undergoes the same sequence of transitions between propositional attitudes as the subject in the original case.⁵

Surely there are striking differences between the two sets of cases. Only in the first set of cases are the subjects responsible, in an agential sense, for their change in view. This explains why only in the original cases would it be appropriate to praise or blame the subject for having formed her view or for the way in which she has formed that view. This difference in agential responsibility is likely connected to a crucial normative difference: for each pair of cases, there are rational norms which only govern the change in view of the original case.

⁵I am inclined to think that for any case of genuine deliberation, we can use the above schema to construct a deviant, non-deliberative version. But even if this isn't right, I see no principled, independently motivated reason to deny that there is at least one possible pair of cases which fit the schema.

Consider the tourist who forms the belief that she is just north of the museum on the basis of her endeavoring to reflect. If she reflects on certain information, and on that basis, transitions to an attitude the content of which conflicts with her reflective information, then she has erred. Not so in the deviant case, for that change in view is arational. This need not preclude us from recognizing norms by which to evaluate the products of both changes in view. Indeed, there is a sense in which, all things considered, it is better to form propositional attitudes with content which is both accurate and supported by evidence. If it is true that the tourist is just north of the museum and the evidence supports a belief to that effect, then in both versions of the case, the tourist's change in view will have culminated in a belief which we may positively evaluate. Note though that an evaluation of the tourist's belief is not an evaluation of the change in view which is responsible for the belief. Only in the original version is the tourist's change in view and its product rationally evaluable.⁶

Similar remarks hold for the other two case pairs, suggesting that there is something normatively significant about an intentional and reflective change in view across the epistemic, practical, and moral domains. A straightforward way to explain the pattern is to treat all three changes in view as belonging to the *sui generis* class of thought of deliberation. Doing so allows us to recognize that in each domain of thought in which deliberation occurs, there will be a subclass of process-norms which govern the deliberative changes in view within that domain. After all, even when we contrast the changes in view of the original cases with a rationally evaluable unreflective, unintentional change in view, there still seems to be an important difference between the two kinds of evaluations. The formation of an unreflective belief formed on the basis of a passive perception might be rationally evaluable, but not in the same way that the deliberative tourist's change in view is evaluable. Again, the explanation would seem to involve the role that the tourist's intentional and reflective agency plays in her change in view.

The contrast drawn by the original and deviant case-sets also underscores the need to treat the intentional and reflective elements of deliberation as causally efficacious. Let us agree that deliberation only occurs in the original cases. In both the original and deviant versions of a case, the subject undergoes the same sequence of attitude transitions. Thus, if we only appeal to the

⁶For a related discussion on the point, see Niko Kolodny's state-requirement/process-requirement distinction (2005).

sequence of attitude-transitions involved in a change in view, we will not be able to explain why only the change in view of the original case qualifies as deliberation. We must think of a change in view as a mental process: a sequence of mental states, along with the causal and functional relations responsible for the sequence. Clearly, there are different causal structures at work in the changes in view of each case-pair. In the deviant version, it is the hallucinatory fever which is the proximate cause of her sequence of attitude-transitions. During deliberation, the subject orchestrates her sequence of attitude transitions, which is just to say that her intentional and reflective agency drives her change in view.

Notice that the point is not just that deliberation is a change in view in which intentional and reflective elements occur. At the outset of the original and deviant version of each case, the subject has an intention to undergo a change in view on the basis of reflection. We may also imagine that in the deviant version, the subject has some kind of conscious awareness of the thoughts entering her mind during the hallucinatory fever.

Nonetheless, the change in view in the deviant version is distinct from deliberation because its intentional and reflective elements are causally inert: despite the clear presence of these elements, they fail to serve the right kind of functional role in the process that is responsible for her sequence of attitude transitions. It is here where the real puzzle which I take deliberation to present arises. How do we explain deliberation as a (person-level) mental process, in which a sequence in propositional attitude-transitions is caused, or otherwise explained by, the deliberator's intentional and rational agency? Put another way, what states or relations must we add to a "bare" sequence of propositional-attitude transitions in order to transform the sequence into an intentional and reflective change in view, that is rightly classified as deliberation? A theory that demystifies the intentional and reflective character of deliberation will respect:

Intentional Reflective Process (IRP) A theory of deliberation should specify, in functional terms, how the intentional and reflective elements of a token process of deliberation explain its sequence of attitude-transitions.

Here is another way of thinking about the problem: what states or relations must we add to a "bare" sequence of propositional-attitude transitions in order to transform the sequence into

an intentional and reflective change in view which deserves to be classified as deliberation?⁷

2.3. The question-directed attitude view of deliberation

One quite natural thought to have is that we could easily have characterized the set of cases in very different terms. We could have said that what is common to each agent is that she is endeavoring to answer some question. Indeed, it is common for philosophers writing on species of deliberation –like theoretical (epistemic), practical, or moral reasoning– to treat it as an activity directed towards questions about what judgment to form or what judgment ought to be formed.⁸ An issue with this kind of proposal is that it is not always clear whether the association between deliberation and questions is metaphorical or intended to stake out a substantive view about the nature of deliberation.⁹ Moreover, it is not clear how the appeal to questions will help an account of deliberation address the Neutrality and IRP constraints.

The view I propose, the Question-Directed Attitude View of Deliberation (QDAV), helps elucidate the ubiquitous, yet nebulous, appeal to questions. According to QDAV, the key to understanding deliberation is to attribute to agents a mental state that is about a question. This will allow us to treat deliberation as a mental process that has the function of answering a question and the sequence of propositional attitude-transitions which occurs during deliberation

⁷Presumably, it won't be feasible to list *all* the necessary states and relations. Some will likely depend on contingent facts about our psychology and biological makeup. Instead, we are looking for the mental states and relations which, once added to a bare sequence of attitude-transitions, will result in a thought process which is intelligible to us as deliberation

⁸To gesture at a few examples, David Owens (2008, p. 261) describes deliberation as a “conscious, intentional activity whose aim is to resolve a certain question”, Stroud (2000, p. 28) thinks that “a philosophical interest in practical reasoning...appears to be more a concern with the role of something called ‘reason’ or reasoning in settling practical questions”, Nomy Arpaly and Timothy Schroeder write that “Theoretical deliberation is primarily concerned with what to believe, while practical deliberation is primarily concerned with what to do” (2012, p. 210), and Nishi Shah writes, “The question that expresses the aim of doxastic deliberation, *whether to believe that p*, immediately gives way to the question *whether p is true*” (2010, p. 248), and “practical deliberation is directed at the question *what to intend* in the sense that it is undertaken with the aim of issuing or not issuing in an intention” (2008, p. 2).

⁹Among those who proffer the more substantive claim, it is rare to encounter an analysis of what it would actually mean for deliberation to be a kind of question-directed thought. As an example, Pamela Hieronymi argues that because “reasoning is (actual or possible) thought directed at some question or conclusion” that “reasons must relate, in the first instance, not to states of mind but to questions or conclusions”(2013, p. 115-6). Despite recognizing a kind of question-directed thought, she distinguishes “states of mind” from questions. If the distinction is meant to exclude mental states that are about questions, then it is unclear what question-directed thought could be. On the other hand, if “state of mind” is meant to refer to a propositional attitude, then perhaps her idea is that reasons must relate to some question which one has a certain attitude towards. If so, we are left with an underdeveloped notion of a question-directed attitude.

as a byproduct of its function. Prior to understanding the mechanics of QDAV and how the view addresses the Neutrality and IRP Constraints, a brief overview of questions as mental content may provide clarity. Section 2.3.1 is devoted to this overview. In Section 2.3.2, I provide a QDAV-style analysis of epistemic deliberation, and I go on in 2.3.3 to explain how the same kind of analysis can be used to capture all deliberation in a way that satisfies the Neutrality and IRP Constraints.

2.3.1 Questions as content

QDAV echoes a proposal by Jane Friedman that there are mental states which are genuinely question-directed (2013). She argues that these states, which she labels the *interrogative attitudes*, form a distinct class of mental state which cannot be subsumed under the umbrella of the propositional attitudes.¹⁰ Instead, such states have questions, not propositions, as content. QDAV contends that deliberation is a mental process that involves an interrogative attitude. I label the interrogative attitude involved in deliberation a *deliberative state* (and I'll continue to reserve "attitude" for the propositional attitudes).

In many respects, recognition of questions as mental content is a natural extension of a well-established practice of treating questions as the semantic contents of an utterance like, "Does the 10:30am train go to Boston?" Because such utterances do not seem to be in the business of communicating anything that could be said to be true or false, the semantic content of such utterances is best thought of as something non-propositional. In a similar vein, QDAV will treat questions as the non-propositional content of the deliberative states which drive the deliberative process. QDAV adopts a partition analysis of questions according to which a question imposes an equivalence relation over logical space, where the resulting partition comprises the set of complete answers to the question.¹¹ In other words, a question is a way of structuring

¹⁰Her general worry is that using propositions to try to model the content of an interrogative attitude will distort the attitude's functional profile. My presentation of QDAV does not explicitly argue that, on functional grounds, the interrogative attitudes are distinct from the propositional attitudes. However, my overall discussion may easily be reframed in such a way as to lend support to Friedman's conclusion. Roughly, the idea is that given the constraints of deliberation presented in Section 2.2, candidate views of deliberation which seek to explain the phenomenon solely in terms of propositional attitudes will not succeed. Given the plausibility of QDAV, we have reason to think that at least one interrogative attitude serves a crucial functional role in explaining some mental phenomena.

¹¹The account of questions we will be adopting draws heavily from Jeroen Groenendijk and Martin Stokhof's view (1984) which elaborates on the core ideas in C.L. Hamblin's (1958) pioneer work on questions. In Groenendijk

possibility into mutually exclusive and exhaustive ways of answering the question:

Q1: Does the 10:30am train go Boston?

Answer Set: ⟨the 10:30am train goes to Boston⟩, ⟨the 10:30am train does not go to Boston⟩

Q2: Who (among Mike, Molly, and Mona) was at the party?

Answer Set: ⟨everyone was at the party⟩, ⟨no one was at the party⟩, ⟨only Mike and Molly were at the party⟩, ⟨only Mike and Mona were at the party⟩, ⟨only Molly and Mona were at the party⟩, ⟨only Mike was at the party⟩, ⟨only Molly was at the party⟩, ⟨only Mona was at the party⟩

Notice that any pair of answers in an answer set are inconsistent. For this reason, the entailment of a complete answer is equivalent to ruling out all other complete answers. Also notice that there also are no propositions we could add to each answer set that would settle the question differently than the answers already included. For these reasons, the answers listed above are said to be complete. (For the sake of brevity, we may ignore tautological questions which only have one complete answer.)

Crucial to the account of deliberation soon to be developed is the notion of a question being *resolved* relative to a body of information. We will say that a body of information (modeled as a subset of logical space) resolves a question just in case it entails a complete answer to that question (equivalently, just in case it is inconsistent with all other complete answers). A body of information partially resolves a question just in case it fails to entail a complete answer but entails the negation of a complete answer (equivalently, it is consistent with at least two answers but inconsistent with at least one answer). Because our notion of question-resolution is relational, a question might be resolved relative to some bodies of information but not others. But, questions are rarely considered in a vacuum devoid of explanatory interests in the kind of body of information that might resolve a question. This is just to say that understanding what

and Stokhof (1984), the intension of a question corresponds to an equivalence relation on logical space. Going forward, I will speak loosely by treating questions as identifiable with an equivalence relation or the resulting partition. For a richer analysis of questions in terms of resolution-conditions see (Ciardelli et al., 2013) and for a decision theoretic, context-dependent analysis of questions see (van Rooy, 2003). Lastly, it should be noted that while a partition analysis of questions may not be adequate for accounting for the full gamut of interrogative utterances (*e.g.*, mention-some questions), I take such issues to be symptomatic of conversational questions and not the questions of the mental contents of deliberators.

it means for a question to be meaningfully resolved for some purpose requires that we have a grip on a *relevant* body of information for resolving the question. One especially pointed way of drawing out this point has to do with the worry that inconsistent information, taken as a set of jointly unsatisfiable propositions, will entail every candidate answer to every question. But, presumably, there will be a number of explanatory interests which will likely suffer if we let inconsistent information qualify as relevant.

As an example, suppose we want to understand the contours of a conversation in which a speaker asks a question. A quite plausible analysis will cite Gricean Maxims and attribute to the speakers the conversational goal of resolving the question.¹² Obviously, not just any body of information which resolves the speaker's question is relevant to the satisfaction of that goal. The conversational goal is met just in case the question is resolved by the Stalnakerian common ground (Stalnaker, 1970) – the shared body of information accepted by the parties of the conversation for the purposes of the conversation. Now, we can begin to understand the various conversational moves between the inquirer and her respondent as a joint pursuit to communicate information by which to update the common ground until it resolves the “question under discussion”.

In a similar fashion, QDAV seeks to analyze deliberation as a mental process in which the deliberator occupies a *deliberative state*: a mental state with a question as content. Just like the aim of a conversational exchange between inquirer and respondent is to resolve the inquirer's question, the function of any token process of deliberation is to resolve the deliberative question which is the content of its deliberative state. To streamline presentation of the view, in the next section, we will focus on the tourist's epistemic deliberation, and how the QDAV strategy addresses the IRP Constraint. In Section 2.3.3, we will abstract away from the particularities of epistemic deliberation to provide a general view of deliberation. In doing so, we will see how QDAV addresses the Neutrality Constraint.

¹²In what follows, I briefly sketch a view in the spirit of Craige Roberts' account (2012). While I cannot do justice to the richness of the view here, it should be noted that the account utilizes the notion of a “question under discussion” to provide a pragmatic analysis for all discourse and not just those conversations in which a speaker asks an explicit question.

2.3.2 An analysis of epistemic deliberation

Suppose that the tourist's epistemic deliberation is governed by a deliberative state that has as content the question *Am I northeast, northwest, southeast, or southwest of the museum?* Whereas a conversational question gets resolved just in case the conversational common ground entails a complete answer to the question, a deliberator relies on her deliberative information to resolve her deliberative question. Deliberative information is the information accessible to her for the purposes of her current deliberation. It consists of the informational contents of whatever attitudes she has that are characteristic of her current deliberative domain. Assume her deliberative information is provided by her various beliefs and doxastic attitudes.¹³

Of course, just because the tourist has certain doxastic attitudes is no guarantee that she will deploy the informational contents of those attitudes during her deliberation. She might be consciously aware that she is near a certain site while having a memory of where the museum is relative to that site. If she never reflectively accesses that memory, its informational content will have no bearing on her deliberative question. The point is that we must take care to distinguish those elements of her deliberative information to which she consciously attends. It is only this information –her conscious deliberative information– which resolves, or partially resolves, her deliberative question by entailing or ruling out a complete answer.

When deliberation commences, the tourist's deliberative question is unresolved, meaning that her conscious deliberative information has not yet determined whether she is northeast, northwest, southeast, or southwest of the museum. Because her deliberation is epistemic, we will assume that she forms the attitude of belief towards complete answers by resolving, or partially resolving, her deliberative question. Perhaps, the easiest way for her to resolve her deliberative question is to gain, and consciously attend to, a new belief about a complete answer *via* testimony or observation. Imagine that in the course of deliberating, she looks up and

¹³Depending on how finely we individuate domains of thought, the tourist's deliberative information might be provided by all of her doxastic attitudes or only those which pertain to local geography. The details of how best to individuate domains of thought will not concern us here. The point is just that once we identify the relevant domain, her deliberative information will be provided by whatever attitudes she has which are characteristic of that domain. Indeed, one's moral judgments will typically not contribute to one's deliberative information during epistemic deliberation as moral information, taken in isolation, tends to be irrelevant to purely descriptive questions. When moral information does seem to be relevant to an epistemic issue, it is usually mediated through some doxastic attitude.

suddenly sees the museum far off to the south. Assuming she consciously attends to this new information, her conscious deliberative information will be updated in such a way as to rule out that she is southeast or southwest of the museum. She will have partially resolved her question, forming the belief that she is northeast or northwest of the museum. If none of the other doxastic attitudes she consciously accesses help determine her location, then she will have partially resolved her deliberative question, forming a belief towards the partial answer *I am northeast or northwest of the museum*. Now imagine that in the course of deliberating, she overhears that the museum is to the southeast of her. In that case, her conscious deliberative information will come to resolve her deliberative question and she will form a belief, the content of which is the complete answer *I am northeast of the museum*.

But often, deliberation is not so straightforward. Many of the cases we care about tend to be more nuanced as they involve elements of “armchair deliberation”. During armchair deliberation the deliberator does not acquire any new information from an external source, yet still makes headway by operating on, integrating, or otherwise manipulating, the information she already possesses. A prime example involves inference in which the deliberator comes to resolve her deliberation by exploiting logical relations between the contents of her information.¹⁴ Imagine that while deliberating, the tourist learns from reading her map that she is by the main entrance of the national park. After a few moments of reflection, she remembers that a local shopkeeper told her that the park’s main entrance is northeast of the museum, and she infers that she is northeast of the museum. That part of her deliberation which occurs after she reads the map qualifies as armchair deliberation because during that period she acquires no new information but still performs some operation which resolves her question.

By demarcating conscious from non-conscious deliberative information, and maintaining that deliberative questions are resolved by the former, QDAV already has the resources to accommodate the above case. The idea is that when the tourist reads the map and learns that she is at the park entrance, her conscious deliberative information comes to include the information *I am at the park entrance*. While her deliberative information includes *the museum is northeast*

¹⁴Perhaps inference is the only way to deliberate from the armchair. But perhaps it is just one form of armchair deliberation. Consider a case of “imaginative” armchair deliberation in which a deliberator deploys information under a conceptually novel circumstance. A thought experiment might elicit an intuition, and on that basis, her deliberative question is resolved. It is unclear whether such a case is best construed as purely inferential.

of the park, she does not consciously attend to that information when reading the map, and thus her deliberative question need not be resolved just because she learns that she is by the park. It is not until some time later when she accesses the relevant memory that her conscious deliberative information will include both *I am at the park entrance* and *the museum is northeast of the park*, allowing her to infer a complete answer to her deliberative question. Crucially, QDAV respects the period of armchair deliberation by allowing a change in conscious deliberative information to occur after the tourist acquires information from an external source (the map).

It is worth noting that QDAV's handling of armchair deliberation is due to its adoption of a kind of fragmentation of the tourist's belief-system. Roughly, belief-fragmentation is the view that the belief-system of an actual agent is best understood as a system of compartmentalized fragments of information, such that i) it is possible for only some fragments to be active at a given time or for a given purpose and ii) it is possible for fragments to contain different, perhaps even conflicting, information.¹⁵ Without implementing some sort of fragmentation on the tourist's belief system, it is unclear how to capture any sort of distinction between those elements of the tourist's information which are active during deliberation and those which are inactive but accessible. It is also unclear how to understand the period of armchair deliberation that resolves her deliberative question. Because her armchair deliberation culminates in the resolution of her deliberative question, her deliberative information undergoes some kind of change during the period of armchair deliberation. However, she acquires no new information during that period, as she already possesses the information acquired from reading the map. Thus, the change in her deliberative information will likely concern a structural change. Absent fragmentation, it is hard to make sense of this.¹⁶

Given that QDAV already presupposes fragmentation in order to maintain the crucial distinction between conscious and non-conscious deliberative information, QDAV may as well help itself to all the advantages afforded by fragmentation. Specifically, QDAV need not maintain that the tourist's deliberative information consists of just two fragments of information, her

¹⁵See (Lewis, 1982; Egan, 2008; Greco, 2014; Stalnaker, 1984; Yalcin, 2016; Cherniak, 1986) for examples of how philosophers have implemented and appealed to belief fragmentation.

¹⁶It is also worth noting that even fragmented accounts will be challenged by armchair deliberation involving necessary truths. A promising strategy is to include in one's fragmentationist framework, a modal resolution on fragments (Yalcin, 2016; Carballo, 2016)

conscious and non-conscious information. Instead the view recognizes that a deliberator's non-conscious deliberative information will often consist of distinct fragments. This allows QDAV to handle more intricate cases of deliberation. For instance, perhaps the tourist might consciously access a fragment corresponding to her episodic memories of local geography without accessing a fragment corresponding to her memories of testimony about local geography.¹⁷

At this juncture, it should be clear that part of QDAV's explanation of the tourist's epistemic deliberation will cite a sequence of transitions between the doxastic attitudes to which she consciously attends during deliberation. We know that deliberation is a mental process – a sequence of mental states along with an underlying causal or functional structure. Moreover, given the constraint IRP, we know that the tourist's deliberation is a mental process in which her attitude transitions can be explained by certain intentional and reflective elements. So, the last piece of the puzzle is to understand how QDAV makes sense of this underlying structure.

Consider first how QDAV can account for the reflective character of deliberation. According to QDAV, the function of a token process of deliberation is to resolve its deliberative question which occurs just in case the deliberator's conscious deliberative information entails a complete answer. Thus, because the tourist's conscious deliberative information factors into the satisfaction criteria of the function of her deliberation, the reflective character of her deliberation can be specified in functional terms. Put another way, the idea is just that the sequence of doxastic attitude-transitions which occurs during her deliberation is the product of her deliberative process pursuing its function. Because satisfying that function requires updating her conscious deliberation so that it entails a complete answer, the various doxastic attitudes to which she attends during deliberation will be functionally explained by an aim to consciously arrive at a complete answer to her deliberative question.

But exactly how is her deliberation intentional? Here, we will again appeal to the function of deliberation. As I have been suggesting, her deliberation can be understood as “aiming” to update her conscious deliberative information in ways that will resolve her deliberative question. Perhaps then, the intentionality of the tourist's deliberation might be captured by attributing to her an intention to update her conscious deliberative information in a way that arrives at

¹⁷Notice that QDAV is not committed to any substantive theory about how agents are fragmented. That, I take it, is a task best left to the cognitive sciences.

a complete answer to her deliberative question. This proposal would capture the sense in which deliberators do not just sit on their laurels hoping to acquire information which, once added to their conscious deliberative information, will entail a complete answer to the deliberative question. During deliberation, we expect the tourist to actively endeavor to bring her deliberation to a close. Of course, this does not mean that she already knows which pieces of deliberative information she must consciously attend to in order to resolve her deliberative question.

Instead, the idea is that the tourist has an intention to consciously attend to doxastic attitudes which, given her present conscious deliberative information, will be more likely to contribute to the resolution of her deliberative question. Suppose that while viewing her map, the deliberating tourist overhears from a reliable source that both she and the museum are in the northern side of the city. Learning that information is sufficient to partially resolve her question. Along with an intention to resolve her deliberative question, what she learns is sufficient to explain why her conscious attention will shift towards that portion of the map representing the northern region of the city. Indeed, had she instead learned that she and the museum were at the southern region of the city, she would consciously attend towards a different region of the map. Such behavior can be explained by attributing to the tourist an intention to resolve her deliberative information, which manifests itself as an intention to consciously attend to information which, once added to her conscious deliberative information, will help resolve her deliberation.

To facilitate the above approach, it may help to adopt a *resource* conception of reflection, according to which conscious attention is best treated as a limited cognitive resource which, when employed by a mental state, contributes to one's overall subjective experience (Peacocke, 1998). Sometimes this is because certain sensations or propositional attitudes demand our conscious attention. Perhaps during her deliberation, the tourist just can't seem to escape the memory of a recent *faux pas*. But at other times, states employ conscious attention because we have directed our conscious attention towards them. If the tourist is concentrating on reading the map, her conscious attention is employed by various map-relevant beliefs and doxastic attitudes. If she hears a loud noise, her attention might drift towards her auditory-states, but with some effort, she could refocus her attention towards her map-related doxastic attitudes. If conscious attention is a resource that the tourist can exert control over, then we now allow for the possibility that she intentionally directs her conscious attention towards doxastic attitudes

which previously contributed only non-conscious deliberative information. As a result, that information becomes conscious and gets included in the information which is responsible for resolving her deliberative question.

While the resource conception of reflection might help explain how the tourist can have an intention to update her conscious information in ways that will resolve her deliberative question, we might worry that the approach over-intellectualizes deliberation. An intention to update one's conscious deliberative information is an intention to consciously attend to certain non-conscious doxastic attitudes. For this reason, an intention to update one's conscious deliberative information is a higher-order propositional attitude –an attitude that is about one's own attitudes. Deliberation might involve such an intention, but if it does, it rules out the possibility that less sophisticated cognitive systems –creatures incapable of having thoughts about their own thoughts– can deliberate. I see no reason why we should reject such a possibility, at least in the absence of additional considerations.

Instead of positing a separate intention to resolve her deliberative question, we might propose capturing the intentionality of the tourist's deliberation within the attitudinal relation of the deliberative state. We will understand a deliberative state as intentional. It is constitutive of occupying a deliberative state that one is disposed to update her conscious information in ways that facilitate answers to her deliberative question. The primary difference¹⁸ between the two proposals is that the current proposal does not posit an attitude that has a content which is about one's deliberative question and state. This would allow us to accept the possibility that an agent incapable of higher-order thought might nonetheless be capable of deliberating on some matter.¹⁹

¹⁸What else might be involved in construing the tourist's deliberative state as intentional? It might involve thinking about her deliberative question as imposing a partition on her conscious deliberative information. In other words, occupying the deliberative state with content *Am I northeast, northwest, southeast, or southwest of the museum?* restructures her conscious deliberative information into classes of information which are consistent with complete answers. The purpose of this cognitive restructuring would be to help her "think" in ways which are relevant to arriving at a complete answer to her deliberative question. She will be cognitively poised to access fragments of her deliberative information which are more likely to help resolve her deliberative question.

¹⁹Perhaps another positive feature of this proposal is that it helps distinguish deliberative states from other interrogative attitudes (states with questions as contents). Consider for instance, the difference in functional profiles between the astrophysicist who is endeavoring to determine whether the number of stars in the universe is even or odd and the casual daydreamer who merely wonders whether the number is even or odd before moving on to a different train of thought. It might well be appropriate to attribute to both an interrogative attitude with the content *is the number of stars in the universe even or odd?*. But presumably only the astrophysicist occupies something like a deliberative state as only she is committed to resolving the content of her interrogative attitude.

2.3.3 The general framework

Section 2.3.2 provided a QDAV analysis of the tourist's epistemic deliberation. It was suggested such an analysis successfully addresses the IRP constraint. What is not yet clear is whether we can abstract away from the particularities of epistemic deliberation and deploy the same sort of analysis in other domains of thought. In other words, it remains to be seen whether QDAV can satisfy the Neutrality constraint. Assuming that the QDAV story of the tourist's deliberation addresses IRP, satisfying Neutrality will secure us with an adequate model of deliberation.

We begin by flagging a fairly modest, yet foundational, assumption about how to represent the information of a given domain of thought in which deliberation occurs. We will then see how the assumption allows us to explain any token process of deliberation in the same way we explained the tourist's deliberation. The assumption: for each domain of thought in which deliberation occurs, there is a model-theoretic way of representing the contents of the kinds of attitudes which individuate that domain. The kind of logical space required to model the informational contents of a domain of thought may, or may not, differ between domains. Perhaps, all content can be modeled in a standard way using sets of tuples constructed out of entities like possible worlds, individuals, and times (Lewis, 1979; Stalnaker, 1984). If so, then there will not be any real difference between the *kind* of content of a non-normative judgment and a normative judgment. What will distinguish one kind of attitude from the other will be the attitudinal relation which a subject bears towards the content of the attitude.

There is a fair amount of consensus that a crucial difference between a non-normative belief and a moral or normative judgment is that the latter possess motivational or prescriptive force (Smith, 1994; Nagel, 1970; Scanlon, 1998; Blackburn, 1984; Gibbard, 2003; Williams, 1979). While some views seek to capture this force in an attitudinal relation to content, others maintain that for at least some normative judgments, the prescriptive force of the judgment is to be located in its content. Put in more general terms, the issue seems to be whether a normative judgment possesses content that is distinctively normative. For views that recognize distinctively normative content, standard representations of content –like sets of world-time pairs– may prove inadequate.²⁰

²⁰An analogous worry also arises for those who seek to reduce normative properties to natural properties while

Fortunately, we can achieve a model-theoretic representation of distinctly normative content by borrowing from Gibbard’s modeling of practical judgments about what to do. To model the content of such judgments, Gibbard posits the abstract entity of a *hyperplan*: a maximally consistent set of plans or decisions for each possible choice situation an agent might face at a possible world (Gibbard, 2003).²¹ He then defines a space of possibilities in which each point is a pair consisting of a possible world and a hyperplan. Because a set of world-hyperplan pairs represents non-normative possibilities along with permissible courses of action for those possibilities, Gibbard takes such sets to be well-suited for modeling the content of a practical judgment. While the details of the proposal might need tweaking, hyperplans provide a promising way of modeling practical content in a distinctively normative fashion. Notice also that we may easily tinker with the apparatus to represent various flavors of normativity. To model normative content of kind N , we merely substitute a hyperplan for an “ N -theory”: a maximally consistent set of binary attributions like $-N$ -permissible or N -impermissible— for each possible-world feature that is relevant to N -normativity. Including N -theories in a construction of logical space should help model content of kind N in a distinctively normative way.

Let us remain non-committal about whether we can use the same construction of logical space to represent the informational content of different domains of thought and about which constructions of logical space are best for representing which domains of thought. We merely assume that if \mathcal{D} is a domain of thought in which deliberation occurs, then there is some construction of logical space $\mathcal{I}^{\mathcal{D}}$ such that the content of any propositional attitude characteristic of \mathcal{D} can be modeled as a subset of $\mathcal{I}^{\mathcal{D}}$. Aside from allowing us to model the information of domain \mathcal{D} , our assumption also lets us define the set of deliberative questions in \mathcal{D} , $\mathcal{Q}^{\mathcal{D}}$, as a set of partitions of $\mathcal{I}^{\mathcal{D}}$. We can now examine the mechanics of whatever deliberation might occur in \mathcal{D} without explicitly referencing the particularities of \mathcal{D} . To do so, we will use $\mathcal{I}^{\mathcal{D}}$ to construct a formal structure by which to model the deliberative information and deliberative

taking normative concepts to be irreducible. Assuming that natural properties are just possible-world features, we might be able to use possible worlds to represent a proposition that is extensionally equivalent to a normative judgment while unable to represent the judgment’s underlying normative concepts.

²¹Gibbard takes a choice situation, or what he calls an “occasion for action”, to be a triple $\langle w, i, t \rangle$ at which an agent i is faced with a set of alternatives for what to do at world w and time t (Gibbard, 2003). A hyperplan p provides a set of practically permissible alternatives for each possible choice situation.

state of whatever deliberation might occur in \mathcal{D} . This will allow us to deploy the same functional analysis of deliberation across all deliberative domains. (See Chapter 3: A Fragmented Information Revision Framework for QDAV for a more detailed walk through)

A deliberative process d occurring in domain \mathcal{D} can be decomposed into the sequence of time slices d_0, d_1, \dots, d_n at which the process occurs. d will be governed by a stable deliberative state, meaning that it persists during the entire sequence. To model the content of this deliberative state, we will use a deliberative question from domain \mathcal{D} , $Q^d \in \mathcal{Q}^{\mathcal{D}}$. Q^d will partition $\mathcal{I}^{\mathcal{D}}$ into a set of complete answers $\mathcal{A}_0, \mathcal{A}_1, \dots, \mathcal{A}_n$, and Q^d will be resolved, or partially resolved, relative to the deliberator's conscious deliberative information. Although Q^d will remain stable during d , the deliberative information of d will be in a state of flux. Hence once we fix Q^d , we are free to think of a time slice d_t in terms of d 's deliberative information at t . The idea then is that the deliberation d can be captured by Q^d and a sequence d_0, d_1, \dots, d_n which represents how the deliberator's information changes during d .

Since we are now treating a time slice d_t as the deliberative information of d at t , it might be tempting to define d_t as a subset of $\mathcal{I}^{\mathcal{D}}$. After all, one's deliberative information is provided by the contents of one's attitudes characteristic of \mathcal{D} . Given that we have defined $\mathcal{I}^{\mathcal{D}}$ as the construction of logical space for which the content of any attitude characteristic of \mathcal{D} can be modeled as a subset of $\mathcal{I}^{\mathcal{D}}$, one might be inclined to treat d_t as a subset of $\mathcal{I}^{\mathcal{D}}$. However, this won't do. In the previous section, we recognized that the tourist's deliberation involved a kind of fragmented belief-system. As a result, it was necessary to demarcate her conscious information from other fragments of her deliberative information. We will extend that insight by treating d_t as a collection of fragments and reserving distinct regions of $\mathcal{I}^{\mathcal{D}}$ to represent distinct fragments.²² We'll use $i_t^d \subseteq \mathcal{I}^{\mathcal{D}}$ to denote the fragment of deliberative information consciously accessed at t and $f \subseteq \mathcal{I}^{\mathcal{D}}$ for a fragment which is accessible at t . We can now define d_t as $\langle i_t^d, F_t^d \rangle$ where F_t^d is the set of fragments accessible at t .

Because $d_t = \langle i_t^d, F_t^d \rangle$, and having already said that d is composed of its deliberative state and the sequence of time slices of its deliberative information, we can express d using Q^d and

²²Notice that because each fragment is a subset of logical space, each fragment of deliberative information will be internally consistent. But, a fragment might contain information which is inconsistent with some other fragment. By construing deliberative information as a collection of fragments, we make it possible to understand those all too real cases in which a deliberator has, within her cognitive reach, conflicting information.

$\langle i_0^d, F_0^d \rangle, \langle i_1^d, F_1^d \rangle, \dots, \langle i_n^d, F_n^d \rangle$. Having modeled the contents of the two core components of a deliberative process, we may now revisit QDAV's proposed functional role for deliberation – to resolve the question of its deliberative state. We know that Q^d is resolved relative to d 's conscious deliberative information so we can state the function of d in more precise terms.

Function of Deliberation: The function of d is to generate a sequence of deliberative information $\langle i_0^d, F_0^d \rangle, \langle i_1^d, F_1^d \rangle, \dots, \langle i_n^d, F_n^d \rangle$ where i_n^d resolves Q^d .

If d has the above function, then so long as d isn't abnormal, d 's total deliberative information should be explainable as an endeavor to update i_t and resolve Q^d . In other words, QDAV predicts that all things being equal, a change between time slices of d 's deliberative information should be interpretable as an attempt to bring one's conscious deliberative information towards resolving Q^d . Also notice that because d 's total deliberative information corresponds to the sequence of propositional attitude-transitions which occurs during d , our analysis of d mirrors the explanation we gave of the tourist's deliberation. In Section 2.3.2, it was suggested that i) we think of her deliberation as an endeavor to resolve a question by consciously accessing the contents of her doxastic attitudes and ii) (i) helps explain the functional role of the intentional and reflective elements of her deliberation as required by the IRP constraint. Thus, our analysis of d should also satisfy IRP.

In addition, notice how the QDAV framework properly addresses Neutrality, which requires that we explain deliberation without explicitly appealing to any feature which is unique to a deliberative domain. By analyzing d 's total deliberative information as an endeavor to resolve Q^d , we made no assumptions about any features particular to \mathcal{D} . The only assumption we made was that there exists some construction of logical space with which to model the information of \mathcal{D} . As a result, the view's commitments are rather sparse and makes no real claims about how best to understand the contours of a particular domain of thought. QDAV is a view of deliberation *tout court*, and is not meant to solve, for example, the issue of whether practical deliberation culminates in a normative belief or a distinct attitude of *intending*.²³ For these reasons, it should be apparent that the view is well-poised to respect Neutrality.

²³For examples of the former view, see (Velleman, 1989; Nagel, 1970; Broome, 2009). For the latter views, see (Bratman, 1987; Harman, 1986; Broome, 2013).

2.4. Closing remarks

To close our discussion, I would like to highlight some additional virtues of QDAV by juxtaposing it with two families of views. Section 2.4.1 discusses views of reflective reasoning or inference in order to draw out a certain kind of non-inferential mental phenomenon. Although this phenomenon falls outside the bounds of reflective reasoning, it nonetheless shares some essential normative features with reasoning. QDAV, I argue, is well-equipped to subsume both reasoning and this non-inferential phenomenon under the umbrella of deliberation, helping explain their shared normative features.

Section 2.4.2 discusses and rejects the idea that reasoning, and other forms of deliberation, can be adequately explained by appeal to higher-order propositional attitudes. On my diagnosis, theorists who are inclined to explain deliberation by way of higher-order propositional attitudes likely do so as in an attempt to satisfy the IRP Constraint (or some similar constraint). I argue that such attempts have failed in the past and will likely continue to fail. In my estimation, this has less to do with the use of a higher-order attitude and more to do with reliance on the propositional attitudes to the exclusion of other kinds of semantic content.

2.4.1 Inferential & non-inferential deliberation

In Section 3.2, we designated the term *armchair deliberation* for a deliberative process in which the deliberator's progress can't be entirely explained by her acquiring information from an external source. A prominent form of armchair deliberation (perhaps the only) is inferential and is often discussed under the headings of inference and theoretical, practical, and moral reasoning. It's the kind of deliberation in which the deliberator experiences a sequence of propositional attitude transitions by intentionally reflecting on the semantic or logical relations between the contents of her attitudes. Let's call such deliberation *reflective reasoning*. The phenomenon of deliberation tends to be discussed much more frequently under the guise of reflective reasoning than in more general terms. I suspect this is due to a widely shared, yet often unstated, assumption that reflective reasoning is the most normatively significant form of deliberation. While reflective reasoning is surely an important object of study in its own right, I want to caution against the assumption by presenting two cases of deliberation which are not

so easily classified as (purely) inferential.

First, it may be helpful to briefly outline two views of reflective reasoning that illustrate a recent approach to the subject (and why talk of reflective reasoning may be subsumed under the umbrella of deliberation). Consider Paul Boghossian's *Taking Condition*, according to which reflective reasoning "necessarily involves the thinker taking his premises to support his conclusion and drawing his conclusion because of that fact"(2014, p. 5). To make sense of what is involved in a thinker *taking* a premise to support a conclusion, Boghossian proposes that the thinker must adopt the conclusion on the basis of her following, or being guided by, a rule. In many respects, the view resembles that of John Broome, according to which reflective reasoning is a mental process in which one constructs the content of a conscious "conclusion attitude" by way of a rule governed operation on conscious "premise attitudes"(2013).

I will not linger on the issue, but it is worth noting that the Boghossian-Broome understanding of reflective reasoning seems to provide inadequate guidance for how to address the IRP Constraint. That constraint requires us to explain deliberation, by specifying the proper causal contributions of the intentional and reflective elements of deliberation. How exactly does a rule-following analysis of reflective reasoning demystify the intentional and reflective elements of reflective reasoning? For Boghossian, rule-following is an unanalyzable primitive. For Broome, a process involves rule-following when the subject has a dispositional attitude of "seeming right" towards the process. When one stands in an attitudinal relation of seeming-right towards a process, one is disposed to lose that attitude upon checking and finding fault with the process (Broome, 2013, p. 238). There is a sense in which such a dispositional state might be able to capture the intentional and reflective elements of reflective reasoning. But does it do so in a way that makes those elements causally relevant to the rest of the process? The causal contribution of a seeming-right attitude occurs when a reasoner checks her reasoning process. But, what if she never checks her process? For all Broome has said, a seeming-right attitude exerts no causal power at the outset of the process. So, even if we agree that a seeming-right attitude might successfully embody some intentional and reflective elements of reasoning, the attitude can only secure a very limited causal role for those elements.

Putting aside such worries, let us turn to the issue of whether the most normatively significant form of deliberation is reflective reasoning. Before we can make progress on that

front, we should get clear on the distinction between reflective reasoning and non-inferential deliberation. Reflective reasoning is deliberation that involves reflecting on the logical or semantic relations between contents and treating certain attitudes as premise-attitudes for some conclusion-attitude. Non-inferential deliberation will just be any deliberation which doesn't qualify as reflective reasoning. The two cases below provide illustrations of non-inferential deliberation.

PERCEPTIVE TOURIST: Similar to TOURIST with the exception that the subject is unable to infer her location relative to the museum by way of the information provided by her map and her memories of the streets and landmarks she recently passed. In an effort to gain further information, she looks up from the map to observe her surroundings. To her surprise, she sees the very museum she wants to visit a short distance away. Immediately, her deliberation about where she is relative to the museum is resolved.

BUSY ATTORNEY: An attorney has the option of filing a certain motion for a particularly nuanced case. She is deeply invested in the outcome of the case and spends considerable time reflecting on the best course of action. However, she only has a limited amount of time to reach a decision, and given the complexities of the case, she remains uncertain about what to do shortly before her deadline. Fortunately, she has a standing policy to turn to a trusted adviser whenever she finds herself in such predicaments. Her adviser –whose legal expertise she greatly prizes– advises her to file the motion. Pressed for time, the attorney immediately forms an intention to file, thereby closing her deliberation. It is not until well after the deadline that the adviser is able to walk her through the intricate legal strategy supporting her advice.

The above cases present as genuine cases of deliberation. Each agent experiences a change in view on the basis of the kind of intentional reflection that is distinctive of deliberation: the

tourist forms a belief about her relative location, and the attorney forms an intention towards her adviser's advice. I also contend that the deliberations which occur in each case are just as normatively significant as reflective reasoning. While normative significance is a murky notion to attribute to a mental process, Section 2.2 provided us with two noteworthy features which seem relevant: the subject of the process is agentially responsible for her change in view, and her change in view is subject to the kind of rational norms which govern intentional, reflective changes in view. Given this characterization, the deliberators in each case seem to be engaged in deliberation that is just as normatively significant as cases of reasoning. Hence, on the condition that their deliberation is non-inferential, the two cases will provide some initial evidence that reasoning might not be the most normatively significant way to deliberate.

What should we make of my contention that the deliberative processes in PERSPECTIVE TOURIST and BUSY ATTORNEY are non-inferential? One might object to the idea that the cases are non-inferential on grounds that each respective agent reasons towards a judgment about whether to seek information or how to seek information, which in turn, helps explain why that agent receives and accepts certain information. But the link between a judgment to seek information and the acceptance of particular information is only causal, not inferential. In other words, reasoning in support of a judgment to seek information is not necessarily reasoning in support of the content of a particular piece of information. Once we make this distinction, it becomes clear that the subjects non-inferentially form certain attitudes that serve a role in their deliberations. If it turns out that they reflect on prior attitudes with an intention to non-inferentially form attitudes to aid deliberation, then we cannot think of their deliberations as reflective reasoning.

Consider the PERCEPTIVE TOURIST. Sometimes, we inferentially form attitudes towards the information provided by our senses. If we are aware that in our present condition a certain perceptual faculty is unreliable, then we may accept the perceptual seemings provided by that faculty as veridical only after submitting them and the faculty to rational scrutiny. But more often than not, we accept the information provided by our senses by default. It is not only possible, but quite likely that the tourist's acceptance of the information she gains by shifting her gaze away from the map is non-inferential. Moreover, it is implausible to treat her judgment to look up from the map as indicative of an intention to gain information which will cause a

reflective, inferential change in attitudes. If this were so, then her intention would not be satisfied by looking up and immediately learning her location relative to the museum without having to reason about her location. If we wish to include her acceptance of that information in her deliberative endeavor, then it is not clear that we can construe her deliberation as purely inferential.

Concerning the BUSY ATTORNEY, a plausible reading of the case may understand her decision to reach out to her adviser as a product of reasoning (as opposed to a product of a well-conditioned habit). Reflecting on the fact that she is at an impasse about an important legal issue and that she has a policy to seek her mentor's advice in such circumstances causes her to realize that her policy has been triggered. Such reasoning does not support the actual content of the advisor's advice. So if her acceptance of the advice is inferential, that is if the attitude she forms towards it is a product of reasoning, it is not due to the reasoning by which she decided to seek her advisor's advice. It was stipulated that she does not understand the advisor's legal reasoning for the advice until after she accepts the advice. Thus, in order for her acceptance to qualify as inferential, the attorney must reason along the following lines:

1. Given my policy, in circumstance *C*, I do whatever-my-adviser-advises
2. I am in circumstance *C*
3. My adviser advised me to file-the-motion
4. Given my policy, I do file-the-motion [1,2,3,]
5. I have no reason not to follow my policy, and whenever I have no reason not to follow it,
I do whatever-I-do-given-my-policy
6. I do file-the-motion [4,5]

Given the description of the case, it seems implausible that the attorney forms an intention to file because she reasons in the above manner. She is pressed for time and would likely accept the advice upon receiving it, bypassing any extended reasoning like the above. More importantly, however, if there is no such thing as non-inferential deliberation, then the only rational way for her to deliberate towards an intention to file the motion would be to engage

in the above reasoning. This would seriously call into question the utility of adopting action-guiding policies. Part of the utility behind forming a policy is to bypass future reasoning about certain issues without penalty to one's status as a rational agent (Bratman, 1987). Presumably, part of the reason why the attorney has the policy she does is to avoid having to reason about accepting the adviser's advice, thereby conserving time and cognitive resources during exigent circumstances. Why, then, can't it be the case that the *only* reflective reasoning which occurs in BUSY ATTORNEY resembles:

1. Given my policy, in circumstance *C*, I do whatever-my-adviser-advises
2. I am in circumstance *C*
3. I do whatever-my-adviser-advises

If this is the only reflective reasoning which occurs in the case, then we allow the busy attorney to accept the advice to file immediately upon receiving that advice. This is just to say that she can form an intention to-file-the-motion solely on the basis of the adviser's assertion and without having to infer.²⁴ If we do not allow her to resolve her deliberation by non-inferentially accepting the advice, then we nullify any advantage to having the policy she does. More generally, if all deliberation is inferential, then it is unclear why planning and adopting policies are rational things to do.

Cases like the ones just discussed invite us to consider the possibility that deliberation includes not just the transitions between propositional attitudes in which deliberators intend to reflect on the logical relations between contents. If some kind of non-inferential act is possible, it is likely to occur during a token process of deliberation which exhibits inferential and non-inferential elements at different phases of the process. A virtue of QDAV is that, should the need arise, it is equipped with resources to recognize three possible kinds of deliberative acts. This affords great explanatory purchase.

²⁴Remember that "reasoning" and "inference" refer to person-level, reflective processes. My comments here should not be read as denying the fact that the busy attorney will be performing many sub-personal inferential processes.

Inferential Act: An act of inferential deliberation occurs when a deliberator directs her conscious attention towards propositional attitudes with the intention that the logical implication of adding those attitudes to her conscious information will resolve her deliberative question.

Non-Inferential Act: An act of non-inferential deliberation occurs when a deliberator directs her conscious attention towards propositional attitudes with the intention that her question be resolved exclusively by those very attitudes.

Promiscuous Act: An act of promiscuous deliberation occurs when a deliberator directs her conscious attention towards propositional attitudes with the intention that doing so will help resolve her deliberative question.

Inferential deliberative acts are easy to wrap our heads around. Think about what the tourist, deliberating about where she is relative to some landmark, does *before* she looks up from her map. Suppose her conscious information includes information about the last intersection she remembers passing while traveling north. While she inspects her map she consciously attends to her “map-attitudes” and does so because she suspects that she will encounter information that, when added to her current conscious information, will allow her to infer her location. If she sees on the map that the last intersection she remembered passing is north of the landmark, then she knows she passed her destination. If it is south of the landmark, then it is possible she has not. She certainly does not intend to find a note on the map which reads, “you are here”.

Perhaps the busy attorney performs an act of non-inferential deliberation when she directs her conscious attention towards learning the adviser’s advice so long as she does so with the intention that those attitudes will be sufficient for resolving her deliberative state. To illustrate a possible act of promiscuous deliberation, consider the tourist when she looks up from the map to observe her surroundings. She does so with the intention to expand her conscious attention to her surroundings in order to resolve her deliberative question irrespective of whether the information she gains inferentially or non-inferentially helps resolve her question. After all, her attempt to expand her conscious information is successful whether she sees a street sign upon looking up from the map or whether she sees the landmark itself. In the former case, her question will be resolved inferentially, and non-inferentially in the latter.

2.4.2 Higher-order views of deliberation

A second prominent strain of thought maintains that deliberation involves “higher-order” mental states –states that are about other mental states one occupies.²⁵ Typically, the higher-order states appealed to are propositional attitudes that are about other propositional attitudes, like beliefs, desires, or intentions about one’s attitudes. For instance, to support the thesis that we enjoy a unique epistemic entitlement towards first-person judgments, Tyler Burge appeals to a picture of reflective reasoning (what he calls “critical reasoning”) which requires that reasoners normally have knowledge about the attitudes operative during reflective reasoning (1996). Similarly, Sydney Shoemaker argues that we have special access to the contents of our mental states, on grounds that without such access –specifically without beliefs and desires about our beliefs and desires– rational deliberation would be impossible. In doing so, he posits the following picture about the states involved in deliberation:

...if the beliefs and desires are all first-order beliefs and desires, i.e., beliefs and desires that are not themselves about the agent’s beliefs and desires, then one thing they do not rationalize is changes in themselves. For such changes to be rationalized the beliefs and desires would have to include second-order beliefs and desires–desires to promote consistency and coherence in the system of beliefs and desires, and beliefs about what changes in the beliefs and desires would be needed in order to satisfy the second-order desires, which in turn would require beliefs about what the current beliefs and desires are. (Shoemaker, 1988, p. 191)

Here, Shoemaker is concerned with epistemic and practical deliberation. He understands epistemic deliberation as producing a change in beliefs and practical deliberation as producing a change in belief-desire sets. The general idea is that in order for deliberation to *rationally* yield a transition between sets of propositional attitudes, A and $A+$, the deliberator must desire that her attitudes have certain attributes, like coherence and consistency, and believe that such attributes demand transitioning to $A+$. This in turn, requires that she have beliefs about what her attitudes are, like the belief that her attitudes are A and that A is incoherent or inconsistent in some particular way (a way which warrants the transition to $A+$ as opposed to some other consistent and coherent set). This kind of picture suggests that rationally evaluating an instance

²⁵There is a sense in which QDAV is a higher-order view because it posits an intention that is about some other mental state, namely a deliberative state. But deliberative states are non-propositional attitudes. The kinds of higher-order views I take issue with in this section are those that appeal to higher-order propositional attitudes that are about some lower-order propositional attitude.

of deliberation would require looking into the higher-order beliefs and desires of the deliberator. Given that deliberation is rationally evaluable, deliberation requires certain higher-order beliefs and desires.

One motivation for adopting a higher-order view is that it may seem to provide the resources required to help make sense of the functional roles of deliberation's intentional and reflective elements (the IRP constraint). Various philosophers have tried to make sense of reflection in terms of a propositional attitude, usually a belief, which is about one's own propositional attitudes. Higher-order states may also provide the resources to capture deliberation's intentional element. If one thinks that a change in attitudes can only be intentional if deliberators occupy a pro-attitude which is in some way about the attitudes involved in the change, then one will be inclined towards a higher-order view. Despite these apparent advantages, the IRP constraint informs us that merely attributing certain higher-order states to the deliberative process will prove insufficient. A proper higher-order account will also need to specify the causal or functional relations between the relevant higher-order states and the propositional attitudes which deliberators transition between. It is here that higher-order accounts of deliberation will encounter difficulty.

To illustrate, let's consider Shoemaker's appeal to higher-order desires about coherence and consistency and higher-order beliefs about what changes are necessary to satisfy that desire. Presumably the move is supposed to parallel Davidson's early view (1963) on which intentionally ϕ -ing is analyzable in terms of a desire and the belief that ϕ is required to satisfy the desire. (Put aside the issues that ultimately led Davidson to abandon this view for one which recognizes a mental state of intention that is not reducible to other propositional attitudes like beliefs and desires (1978).) We still encounter a problem when trying to explain intentionally ϕ -ing when ϕ is a mental act like transitioning between propositional attitudes. It seems obvious that an agent may desire that her beliefs are true, believe that her desire is more likely to be satisfied if she avoids rushing to judgment about issues, while still rushing to judgment about a great many issues. Similarly, it is not at all clear how the desire to occupy consistent and coherent attitudes along with beliefs about what changes in attitudes are necessary to satisfy that desire can directly cause a changes in attitudes. This is problematic because deliberation is a process involving a transition between attitudes, often concluding in an immediate change in attitudes.

Notice that the issue is unrelated to the fact that we have been trying to explain an intentional act (a deliberative change in attitudes) without reference to an intention. Suppose we attribute to deliberators an intention to occupy a set of attitudes which are consistent and coherent. Ordinarily, there is no great mystery how an intention to ϕ immediately gives rise to ϕ when ϕ is a basic act; it is usually sufficient to explain the raising of my hand by citing my intention to do so. But things change when we consider non-basic acts –like cooking a meal or securing the promotion– the successful performance of which is mediated by a collection of basic acts. Some mental acts seem basic. Intending to imagine or pretend seems sufficient for imagining or pretending. But an intention to occupy a consistent and coherent set of attitudes is certainly not sufficient for bringing about such a change in one's first-order propositional attitudes. The trouble though is that the intentional aspect of deliberation seems capable of causing a direct and immediate change in one's attitudes.

As an illustration, suppose I intend to believe that I will receive the promotion in order to satisfy my desire to diminish my anxiety. Perhaps my intention can explain why I intentionally place myself in situations in which I believe myself likely to encounter evidence in support of my receiving the promotion and intentionally avoid situations in which I believe myself likely to encounter evidence to the contrary. But a majority of philosophers would deny that my intention can directly cause me to form the belief that I will receive the promotion in the same way that an intention to raise my hand can directly cause the raising of my hand (the few direct doxastic voluntarists are the notable exception). Analogously, it would seem that an intention to change one's attitudes to satisfy some desire about those attitudes cannot directly cause that change in attitudes. At best, it can cause future events which raise the likelihood that the change in attitudes will occur. But deliberation often terminates in an immediate change in attitudes, not a possible means for how to bring about that change. This does not mean that no higher-order view can succeed in explaining deliberation. It just means that a higher-order view must be able to specify the right functional relations between the higher-order states in order to satisfy IRP. Here is another way to frame the issue. In order to address IRP, higher-order views must address the *direct descent problem*: how does a higher-order state(s) directly cause or contribute to a change in some lower-order propositional attitudes?

In my estimation, the descent problem is only an issue for higher-order views which attribute to deliberation higher-order states that are about propositional attitudes. QDAV avoids the problem by positing a higher-order state that is about a non-propositional attitude, namely, an intention to resolve one's deliberative state. An intention to resolve a deliberative state just is an intention to direct one's conscious attention in ways that will help resolve one's deliberative question. By resolving a deliberative question, even just partially, one comes to form a propositional attitude towards one of the question's answers. So if one accepts that we have some intentional control over conscious attention, and we resolve deliberative questions via conscious attention, then it becomes clear how our intentions can play a direct role in bringing about an attitude without directly influencing its particular content.

Chapter 3

A Fragmented Information Revision Framework for QDAV

3.1. Introduction

According to the Question-Directed Attitude View (QDAV) of deliberation presented in “Deliberations on Deliberation?” the mental process of deliberation involves a deliberative state, the content of which is a question, and the function of the process is to resolve that deliberative question. Our purpose here will be to explore some of the normative upshots of QDAV. To do so, Part I will provide a formal framework for how to think about QDAV. Although the formal implementation will make some idealizing assumptions (as all models must), it should provide a useful starting point for rigorously exploring the normative implications of QDAV which we finally do in Part II. Thus, Part I should be understood as offering a descriptive framework, while Part II engages some of the normative upshots of that framework.

We will begin in Part I, Section 1 by introducing a formal implementation of QDAV which naively assumes that a deliberator’s information can be represented by a single region of logical space. In Section 2, we will abandon this naive implementation for a fragmented implementation of QDAV (which more closely aligns with the presentation of QDAV in “Deliberations on Deliberation”). The picture that emerges envisions a token instance of deliberation as constituted by a sequence of updates to a deliberator’s conscious information along with a deliberative information structure. Such structures fix the information fragments which are accessible to the deliberator during her deliberation, and as such, are crucial for explaining the sequence of updates to the deliberator’s conscious information.

An important consequence of the picture developed is that each update to a deliberator’s conscious information will be the product of two kinds of mental operations: the deliberator’s consciously accessing a previously non-accessed fragment of information (selection), and the

integration of the newly accessed fragment into the deliberator’s conscious information (integration). Part I provides formal constraints for the integration operation. These constraints should be treated as descriptive; they detail how, under normal conditions, a deliberator’s conscious information will be updated upon integrating a selected information fragment. By contrast, no such constraints are provided for the selection operation. This is because that operation—the directing of one’s conscious attention to particular information—is taken to reflect the deliberator’s agency.

In Part II, we will explore the normativity of deliberation *via* two routes. Part II, Section 1 discusses some candidate norms for the selection operation. There, the presumption is that a given instance of deliberation is rationally assessable in virtue of norms governing the kind of information a deliberator ought to selectively attend to during deliberation. In Section 2, we adopt a virtue theoretic approach. Instead of focusing on evaluations of some token deliberative process, we will look at ways to evaluate the information structures by which one deliberates.

Part I

3.2. The naive implementation of QDAV

How might we begin to think about how to represent a token deliberative process on the QDAV framework? We know that for the duration of any deliberative process, the deliberator will occupy a deliberative state, Q^d , the content of which is a deliberative question. Following the proposal presented in “Deliberations on Deliberation?”, deliberative questions are understood as partitioning whatever construction of logical space is used to represent the informational contents of a domain of thought. If we let $\mathcal{I}^{\mathcal{D}}$ be whatever construction of logical space such that any informational content of domain \mathcal{D} is representable as a subset of $\mathcal{I}^{\mathcal{D}}$, then a deliberative question in \mathcal{D} will be representable as the equivalent class of $\mathcal{I}^{\mathcal{D}}$ corresponding to the set of complete answers to the question.

To make matters a little more concrete, consider a detective deliberating about who committed a certain crime. Certain that one, and only one, of four suspects committed a crime, her deliberative state can be modeled as $Q^d =$

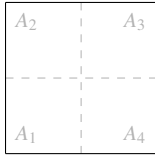
A_1 : (only) the banker committed the crime

A_2 : (only) the lawyer committed the crime

A_3 : (only) the doctor committed the crime

A_4 : (only) the politician committed the crime

And we can visually portray the content of Q^d as:



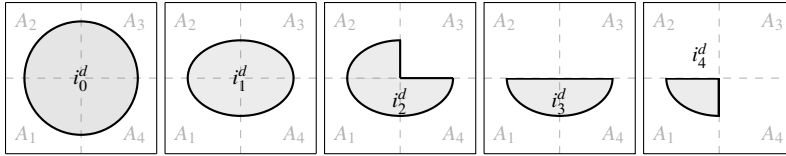
Given that her deliberative state is what guides her deliberation, she will occupy her deliberative state during the entirety of her deliberation.

Next, we will want to represent a deliberator's *conscious information*. Her conscious information is just whatever deliberative information she has consciously or reflectively accessed for the purposes of deliberation. If she deliberates in domain D , and I^D can be used to represent the informational contents of D , then we should be able to use \mathcal{I}^D to represent her conscious information. Eventually, we will be forced to complicate the proposal, but for now, let's use a (non-empty) subset of \mathcal{I}^D to represent an agent's conscious information. Because her conscious information will be in a state of flux during deliberation, we will do well to use the region $i_t \subseteq \mathcal{I}^D$ to represent her conscious information at time t . Doing so lets us represent her deliberative process d as a deliberative state along with a sequence of snapshots of her conscious information. Let's call this the Naive Implementation.

Naive Implementation: An agent's deliberation d , occurring from $t_0 - t_n$ and governed by a deliberative state Q , will be modeled as $d = \langle Q, \langle d_0, d_1, \dots, d_n \rangle \rangle$, where each time slice $d_t = i_t$

On the Naive Implementation the detective's deliberation will look something like this:

When her deliberation initiates at t_0 , her conscious information has not yet ruled out any answers to her deliberative question (for every A , $i_0 \cap A \neq \emptyset$). The function of deliberation is to

Figure 3.1: d during $t_0 - t_4$

resolve the deliberative question which occurs just in case her conscious information entails an answer (*i.e.*, is inconsistent with all other answers). This occurs at the last time slice t_4 , as i_4 is only consistent with A_1 , bringing her deliberation to a close. Suppose, on the other hand, that her deliberation terminated at t_3 . Were that the case, her conscious deliberative information will be consistent with some answers (A_1 , A_2 , and A_4) but inconsistent with others (A_3), meaning that she has only partially resolved her deliberative question. Of course a deliberative process which partially resolves a question is typically preferable to one which fails to rule out any answers. Were her deliberation to have terminated at t_1 , she would fail to have even partially resolved her question as i_1 is still consistent with $A_1 - A_4$.

We won't speculate as to why a deliberative process might end without resolution, or partial resolution, of a deliberative question. We will take it as given that a process of deliberation might end without fulfilling its function and that there is some way of identifying the time slice at which an unsuccessful process of deliberation draws to a close. Perhaps, another truism is that from the point of view of the deliberator, it is best to come as close as possible to resolving the deliberative question. Here the idea is just that genuine deliberation requires of the deliberator that the only qualitative difference she recognizes among answers is *correctness*. But at the outset of deliberation, she has not yet determined which answers are correct and which are incorrect – that's what her deliberation is all about. Given that all answers are to be treated equally until some are ruled out, the deliberator should prefer for her deliberation to come as close as possible to resolving her deliberative question. Call this the *Closeness Norm*.

Notice that because the set of answers to a deliberative question are mutually exhaustive and exclusive, the maximum number of answers which can be consistently ruled out will always be one less than the total number of answers to the deliberative question. Insofar as ruling out the maximum number of answers to a deliberative question just is the resolution of the deliberative question, the Closeness Norm might seem to embody the function of the deliberative process.

But, as we will see later it is unclear how best to implement the Closeness Norm, especially when there is uncertainty about when the deliberative process will terminate.

3.2.1 Problematizing the naive implementation

While the Naive Implementation helps capture some elements of deliberation, it leaves out some rather essential information. When a deliberator updates her conscious information, she does so by integrating into her conscious information some information which previously escaped her conscious attention. Although the Naive Implementation can represent a deliberator's conscious information, it has no mechanism for representing the information which a deliberator has accessed in order to integrate into her conscious information. For this reason, the Naive Implementation lacks the resources to explain how it is that a deliberator's conscious information gets updated from one time slice to the next.

This shortcoming of the Naive Implementation is symptomatic of a deeper issue. The view doesn't just lack a way of representing the information which a deliberator uses to update her conscious information. It also lacks a way of representing the information which a deliberator could have used to update her conscious information. After all, it is often the case that a deliberator has reflective access to various kinds of information. Assuming that how she ought to update her conscious information depends on the information which is accessible to her, the Naive Implementation does not equip us to make certain rational evaluations of a deliberative process.

To make the point concrete, let's specify the deliberative process of the detective depicted in Figure 1 and then contrast it with that of a second detective.

A_1 : (only) the banker committed the crime

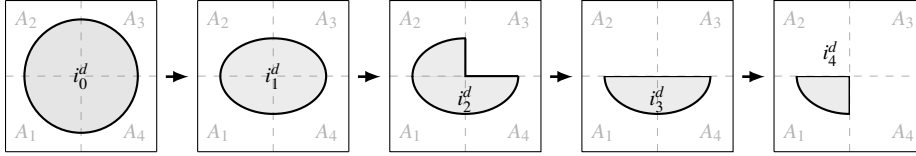
A_2 : (only) the lawyer committed the crime

A_3 : (only) the doctor committed the crime

A_4 : (only) the politician committed the crime

Detective 1 begins by consciously attending to (1) and (2) includes:

1. *The suspect fled with the stolen goods*

Figure 3.2: Deliberation d of Detective 1

2. *If the escape route was a tunnel, then the suspect was under 6 feet*

Although she has learned (3), she has not yet accessed that information at t_0 .

3. *The escape route was a tunnel*

Between $t_0 - t_1$, she reflects on (3), thereby integrating it into her conscious information i_0^d . The result is i_1^d which does not rule out any suspects per se but does require that the guilty party is 6' or shorter. While she has not yet reflected on the following information, she also has access to (4) - (6).

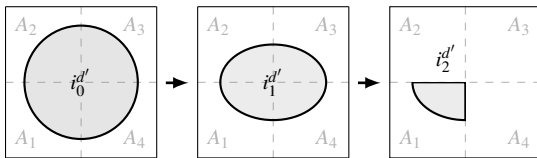
4. *The doctor is not under 6'*

5. *The lawyer is not under 6'*

6. *The politician is not under 6'*

Detective 1 accesses and integrates (4) into her conscious information during $t_1 - t_2$, (5) during $t_2 - t_3$, and (6) during $t_3 - t_4$. The corresponding result is that i_2^d rules out the doctor having committed the crime, i_3^d rules out the doctor and lawyer having committed the crime, while i_4^d settles on the politician having committed the crime.

Now consider Detective 2 who deliberates over the same issue.

Figure 3.3: Deliberation d' of Detective 2

Notice that $i_0^d = i_0^{d'}$, meaning that Detective 2 begins by consciously attending to the same information as Detective 1. We'll also stipulate that both Detectives reflect on the same information during $t_0 - t_1$. So at t_1 , Detective 2's information will also include (1) - (3).

1. *The suspect fled with the stolen goods*
2. *If the escape route was a tunnel, then the suspect was under 6'*
3. *The escape route was a tunnel*

Like Detective 1, Detective 2 also has access to information about the height of each suspect. But instead of accessing (4) - (6), Detective 2 deliberates differently. She has access to (7) which she reflects on during $t_1 - t_2$.

7. *If the suspect fled with the stolen goods, and the escape route was a tunnel, then the banker committed the crime*

Accessing (7) results in Detective 2 arriving at the same answer to her deliberative question as Detective 1. But, Detective 2 took a quicker route to that answer. Assume that the banker did commit the crime and that neither detective irrationally acquired any of their accessible information. Whose deliberation is more rational? If efficiency is a rational consideration, then perhaps Detective 2's deliberation is the more rational. Detective 2 would certainly be criticizable had she performed d instead of d' . Notice though, that it would be premature to judge that Detective 1's deliberation was irrational. While we stipulated that Detective 2 had reflective access to all of Detective 1's information, we did not specify whether Detective 1 had access to all of Detective 2's information. If Detective 1 lacked access to (7), then she could not have performed d' . Perhaps, d was Detective 1's only route to determining that the banker committed the crime. In which case it would be improper to fault her for performing d' . The lesson, then, is that certain rational evaluations will depend on what information is accessible during deliberation. Assuming it is worthwhile to make evaluations about which pieces of accessible information a deliberator should attend to during deliberation, the present way of representing a deliberative process will not do. To this end, we will develop the Fragmented Implementation of deliberation.

3.3. The fragmented implementation of QDAV

Like the Naive Implementation, the Fragmented Implementation will continue to use a stable deliberative state, Q to model deliberation. Unlike the Naive Implementation, the Fragmented

Implementation won't just be sensitive to the information that a deliberator consciously attends to. To include all the information which is accessible to a deliberator, we will represent her information as a collection of compartmentalized fragments, where each fragment is reflectively accessible – meaning that she can consciously attend to the fragment and integrate it into her conscious information. We will use $f \subseteq I^D$ for an information fragment that is accessible to a deliberator and will continue to use i for that fragment which she reflectively accesses. (Notice that we are making the idealizing assumption that each fragment of deliberative information is consistent and closed under entailment. However, we allow that a deliberator can have inconsistent fragments.) A time slice of deliberation will now resemble something like the following:

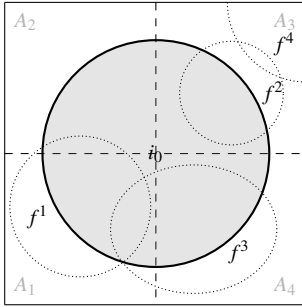


Figure 3.4: a time slice representing information fragments

How should we represent the next time slice of the deliberative process? Let's keep matters simple and assume that during the entire deliberative process the only element of the deliberator's information which is subject to change is her conscious information i_t . In other words, if $F = \{f^1, f^2, f^3, f^4\}$ is the collection of accessible fragments that the deliberator is not attending to at t_0 , then her total information at any at any time t during deliberation will be comprised of $i_t \cup F$. Such an assumption is, of course, false. During deliberation, an agent might acquire new deliberative information in any number of ways: reading an article, hearing testimony, or witnessing an event. It is also true that during deliberation, an information fragment might be destroyed or altered as a result of the deliberator accessing that fragment. Suppose that while deliberating about the Mueller investigation, I expand my attention to include an information fragment comprised of my judgments about the veracity of various news sources. The contents of that fragment might very well be revised if my conscious information already includes

information which is relevant to the veracity of some of those news sources. However, our assumption allows us to focus on how the accessible information possessed by an agent can shape her conscious information.

Although F won't change at the next time slice, what of her conscious information? After deliberation commences, a deliberator's conscious information is just the product of her having updated her conscious information at a prior time with a fragment of accessible information. So, her conscious information at t_1 should depend on which fragment she accesses during $t_0 - t_1$. If we use \bullet to signify the operation of integrating an accessible information fragment into a deliberator's conscious information, then for whatever $f \in i_0 \cup F$ the deliberator accesses during $t_0 - t_1$, $i_1 = i_0 \bullet f$. To indicate on our diagrams which fragment a deliberator accesses, we will use a dotted region. To illustrate, let's take an easy case in which a deliberator fails to access a new fragment of information. We'll capture this by representing her as "selecting" i_0 to (re)access during $t_0 - t_1$. Intuitively, this would seem to yield:

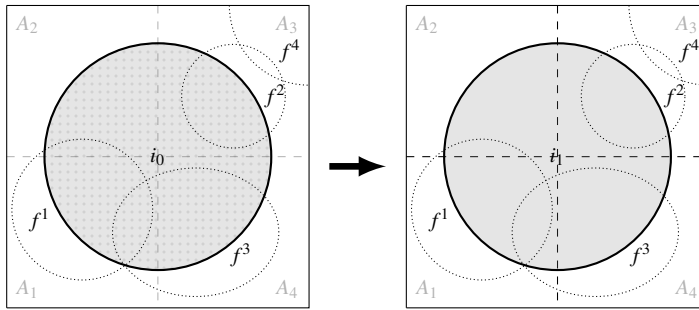


Figure 3.5: $i_0 \bullet i_0 = i_1$ (failure to access a new information fragment)

Now suppose that during the next juncture of the process, the deliberator accesses a fragment of information which she has not yet accessed. Specifically, suppose she updates with f^1 during $t_1 - t_2$:

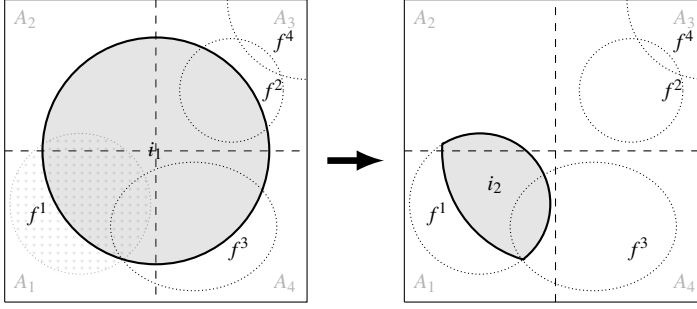


Figure 3.6: $i_1 \bullet f^1 = i_2$ (updating with f^1 during $t_1 - t_2$)

Because some of the information the deliberator integrates into her conscious information during $t_1 - t_2$ overlaps with i_1 , the intuitive result of such an update would be to take the intersection of i_1 and f^1 . Here the idea is just that if one already has the information *someone from the Mueller investigation will leak details about the report* and *Congress might gain access to details about the report*, then one does not lose the information *someone from the Mueller investigation will leak details about the report* just because one learns that Congress will gain access to details about the report. In other words, whenever information is integrated into a deliberator's conscious information, the deliberator will continue to consciously attend to whatever information she has already accessed which is consistent with her new information. Clearly, this is an idealization. But by and large, I take the idealization to describe the way in which deliberators update the information which they use to resolve deliberative questions.

We are now well poised to summarize the Fragmented Implementation of QDAV. As on the Naive Implementation of QDAV, we are still committed to understanding deliberation as a sequence of time slices d_0, d_1, \dots, d_n aimed at resolving a deliberative state Q^d . After the first time slice d_0 , we will want to analyze each time slice as the product of updating the deliberator's conscious information at the prior time slice with some accessible information fragment. To do this, we introduced $F = \{f^1, f^2, \dots, f^n\}$, the collection of information fragments which will be accessible to the deliberator during d . Doing so lets us think of d_1 as $d_1 = i_1 = i_0 \bullet i$ for some $i \in i_0 \cup F$. Subsequent time slices will be analyzed in the same manner. Notice though that the strategy requires that we have values for i_0 and F .

To fix i_0 and F , we will introduce the notion of a *deliberative information structure*, $\langle Q^d, i_0, F \rangle$,

which we can think of as providing all of the deliberator's information which is relevant for analyzing the sequence of time slices of d . i_0 will be supplied by various contingent facts about the deliberator when she first starts deliberating about Q^d , while the function \mathcal{F} which will fix her accessible information fragments during d . We can think of \mathcal{F} as supplied by facts about the deliberator's history in conjunction with her cognitive architecture – for any deliberative question Q , $\mathcal{F}(Q) = F^{d*}$ where F^{d*} is the collection of accessible fragments the deliberator will have during any possible deliberation about Q . Notice that this allows the flexibility to recognize that some information fragments might be accessible to one when deliberating about certain questions but not others.

The idea then is that on a fragmented implementation of QDAV, some token deliberation d will consist of a sequence of time slices and an information structure. Out of convenience, we will sometimes talk about a deliberative process, meaning some token deliberation, as if it were merely a sequence of updates to a deliberator's conscious information. When we do, it should be clear from the context what the relevant information structure for the deliberative process will look like. Below, the fragmented implementation of QDAV is summarized in more detail.

Fragmented Implementation (first pass): Deliberative process d will be modeled using $\langle \mathcal{I}^D, \mathcal{Q}, T, \bullet \rangle$.

- where D is the deliberative domain of d , \mathcal{I}^D will be the appropriate set-theoretic construction of possibilities needed to capture the information of domain D
- \mathcal{Q} is the set of equivalent relations over \mathcal{I}^D corresponding to the set of all possible deliberative questions in domain D .
- \mathcal{T} will be a linearly ordered set of time instants such that for some $\langle t_0, t_n \rangle \in \mathcal{T}$ (where $0 < n$), d_0 is the initial time slice of d and d_n is the last.¹
- where $i \in \mathcal{I}^D$ is information a deliberator consciously attends to and $f \in \mathcal{I}^D$ a fragment of accessible information, $i \bullet f$ is the result of integrating f into i .

¹Though it is convenient to do so, we need not assume that d always starts at t_0 . We could let d begin at t_m and end at t_n where $0 \leq m < n$. This would let us model two possible deliberations which begin at different times. In Part II, we will be interested in using the model under development to make comparative evaluations about different possible deliberations. So, one might worry that because our model requires that deliberation starts at t_0 , the only kinds of evaluations supported by the model will be restricted deliberative processes which all start at the same time. Fortunately, the kinds of evaluations we will be considering will look at the efficiency of a deliberative process in which a given start and end times of the process are irrelevant (although the length or distance between start and end times *will* matter). Nothing in the discussions to follow in Part II will be lost by assuming that the deliberative processes we wish to evaluate all start at t_0 .

d will consist of an information structure $\langle Q^d, i_0, \mathcal{F} \rangle$ and the sequence d_0, d_1, \dots, d_n . $Q^d \in \mathcal{Q}$ is the deliberative state governing d , $i_0 \in \mathcal{I}^D$ is the deliberator's conscious information at t_0 , and $\mathcal{F}(Q^d) = \{f^1, f^2, \dots, f^n\}$ is the collection of information fragments accessible during every d_t . For each d_t :

1. if $t = 0$, then $d_t = i_0$
2. if $t \neq 0$, then $d_t = i_t = i_{t-1} \bullet i$ for some selected fragment $i \in i_{t-1} \cup \mathcal{F}(Q^d)$,

In Section 3.1, we'll make an important modification to the Fragmented Implementation. But before doing so, it is worth highlighting an important feature of the view: with the exception of the first time slice, the view models a time slice of deliberation as an update on the prior time slice. Notice also that each update consists of two operations: the *selection* of an accessible fragment of information and the *integration* of that fragment into a deliberator's conscious information. For this reason, we can identify two potential dimensions by which to evaluate a deliberative process. Selection Norms evaluate whether a deliberator accessed the best fragment of information to integrate into her conscious information. Integration Norms help us evaluate whether the deliberator acted rationally in modifying her conscious information in light of the fragment of information which she accesses. Eventually, I will suggest that only Selection Norms are genuinely normative. Integration Norms are best thought of as constitutive constraints on the integration operator \bullet .

Distinguishing between Selection Norms and Integration Norms provides for a useful reframing of a shortcoming of the Naive Implementation. Recall that the Naive Implementation failed to support evaluations of a deliberative process based on what information is accessible to a deliberator. This is really a failure to recognize norms which govern the kind of accessible information a deliberator ought to reflectively access during deliberation. Because the Fragmented Implementation provides for explicit representations of a deliberator's accessible information, it allows for the investigation of Selection Norms. As an example, consider a deliberator with the following information structure.

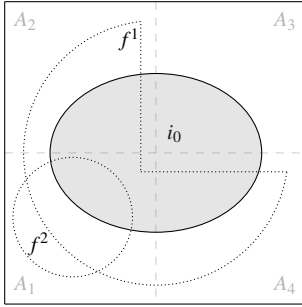
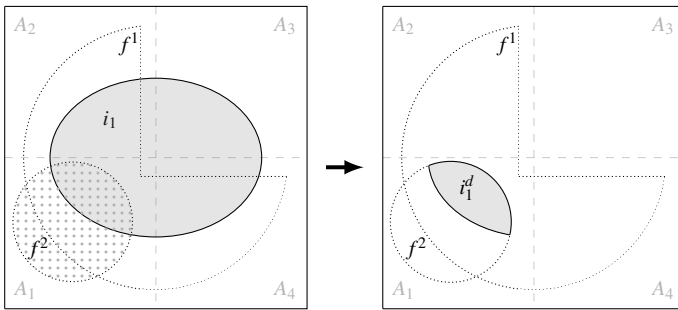
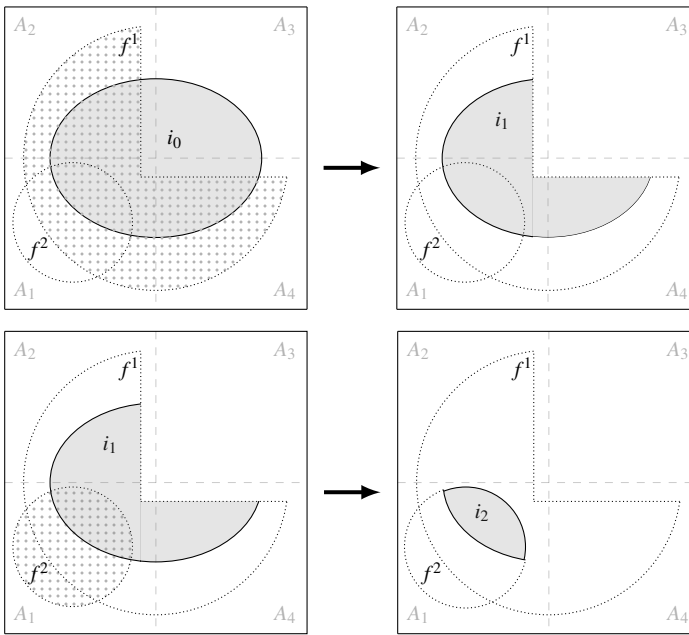


Figure 3.7: Deliberative Information Structure STR

Given her accessible information, fragments f^1 and f^2 , there are really just two ways to proceed from STR and arrive at A_1 : $d^1 = \langle i_0^{d^1} \bullet f^2 \rangle$ and $d^2 = \langle i_0^{d^2} \bullet f^1, i_1^{d^2} \bullet f^2 \rangle$

Figure 3.8: deliberative process d^1 Figure 3.9: deliberative process d^2

Assuming that A_1 is the correct answer, d^1 is clearly the more efficient means of deliberating. If one thought that the greater efficiency of d^1 rendered the process the more rational of the two, then it would be desirable to have a model of deliberation which offers resources for analyzing one's evaluation. For instance, we might entertain a norm that at t requires a deliberator to update her conscious information with that accessible fragment of information which will bring her closest to resolving her deliberative question. Note that my point here is not to promote a particular Selection Norm. I only mean to illustrate how the Fragmented Implementation allows the articulation of a Selection Norm which, in turn, can help us make sense of certain evaluations of a deliberative process.

We will be treating Integration Norms as partially constitutive of deliberation (and conscious thought more generally). I take the integration of an accessible fragment of information into a deliberator's conscious information to be a largely passive operation. If you're currently reflecting on *the banker or the politician did it*, and you consciously attend to *the politician didn't do it*, then you're rationally required to update your conscious information in way that is consistent with the banker having done it. Under normal circumstances, an agent will satisfy that requirement without much effort. From the agent's perspective, it will likely seem as if her thinking that the banker did it was a passive effect of her thinking that the politician didn't do it. A plausible explanation is that Integration Norms are constitutive of reflective thought: a pattern of failing to reasonably integrate new information into one's conscious information suggests a general inability to deliberate. By contrast, an agent can have a pattern of violating Selection Norms without relinquishing her status as a deliberator. She will merely qualify as a poor deliberator.

In light of the above, I will take the identification of Integration Norms to be relatively more straightforward than the identification of Selection Norms. Section 3.1 presents Integration Norms in the guise of formal constraints on the integration operation. To do so, we will draw on some insights from the belief revision literature. That literature is predominantly concerned with formally characterizing the minimal changes required for a belief state to rationally accommodate some incoming piece of information. Although the deliberative framework we are after is not meant to be restricted to transitions between belief states, a similar issue arises for our framework: what are the minimal changes required for a deliberator's conscious

information to rationally integrate some fragment of information? It might be that the correct response diverges significantly from the rational requirements on belief revision. Nonetheless, it is as good a place as any to start.

3.3.1 Fragmented implementation: second pass

The current version of the Fragmented Implementation requires us to think of a deliberator's conscious information i_t as $i_{t-1} \bullet i$. Unfortunately, this way of thinking about the deliberative process is insufficient. To see why, consider what the deliberative process depicted below should look like at t_2 .

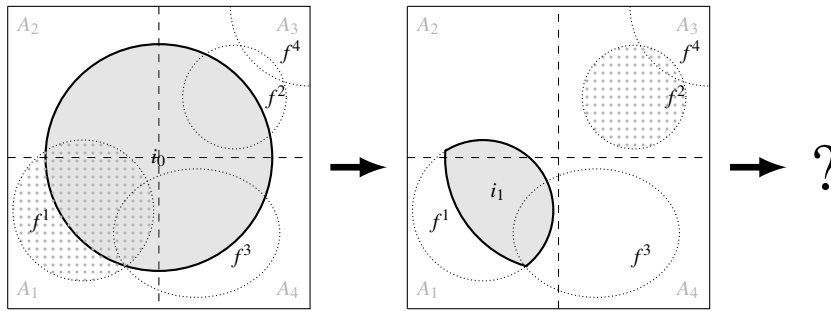


Figure 3.10: What should $i_1 \bullet f^2$ look like?

Given the current framework, $i_2 = i_1 \bullet f^2$. But f^2 does not intersect i_1 , so how are we to understand $i_1 \bullet f^2$? During $t_0 - t_1$ and $t_1 - t_2$, the fragment to be integrated was consistent with the deliberator's pre-existing conscious information. As a result, the product of those integration operations was just the intersection of her conscious information and the fragment being integrated. But because i_1 and f^2 are inconsistent, treating $i_1 \bullet f^2$ as their intersection would force us to treat i_2 as empty. That's clearly not desirable. Our model should allow a deliberator to productively update her conscious information with information that is inconsistent with her conscious information. Suppose that according to a deliberator's conscious information, *someone from the Mueller investigation will leak* and *Putin is the puppet master*. If the deliberator then consciously attends to her memory of having learned that no such leak will occur, she can still retain her conscious information about Putin along with whatever other information is unrelated to whether or not there will be a leak. This is because her conscious information will come to accommodate her memory in light of conditional information about how things would

be were she to learn that no leak will occur.

The lesson, then, is that we need to be sensitive to certain conditional information dictating the counterfactual possibilities recognized by a deliberator's conscious information. Because such conditional information is not encoded by a mere region of logical space like i_t , we need to revise how we are thinking about conscious information and the integration operation. We will say that a deliberator's conscious information at t will consist of i_t along with an associated system of spheres \mathcal{S}_t centered on i_t . Envision \mathcal{S}_t as akin to a system of Lewisian spheres (1973) representing conditional information. The key difference will be that while Lewisian spheres are centered on a single point in logical space, our sphere system will be centered on the set of points i_t . Those spheres which are closer to i_t –the smaller spheres– represent possibilities which are closer to how things are according to the deliberator's conscious information. Figure 11 provides an example of i_1 with an associated system of spheres.

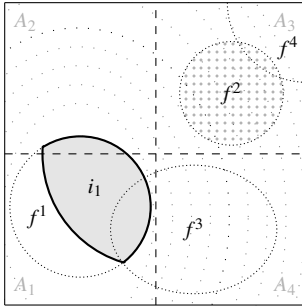


Figure 3.11: i_1 & \mathcal{S}_1

The boundaries of each sphere are represented by loosely dotted lines. Interpret the space between boundaries as containing possibilities which are equally close to i_1 . i_1 's system of spheres is useful because it provides a way of thinking about how the deliberator's conscious information will look after integrating with f^2 . The general idea is that the result of integrating f^2 will be the smallest region of f^2 which is contained in one of \mathcal{S}_1 's spheres. This captures the sense in which consciously accessing and integrating some fragment of information should only require accommodation by way of minimal changes to the deliberator's conscious information. Consider Figure 12.

Here's another way of putting the point. i_1 's system of spheres provides a way of ordering possibilities depending on how close those points are to the deliberator's conscious information

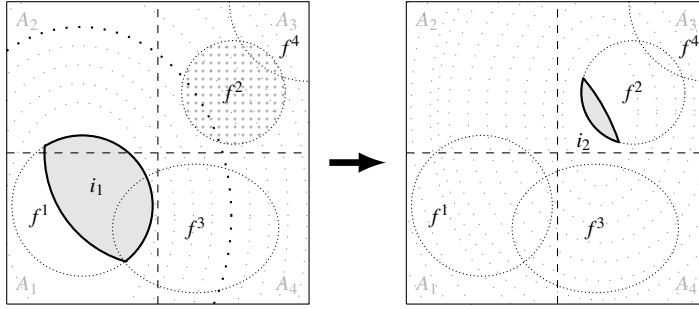


Figure 3.12: updating i_1 with f^2 looks at the smallest sphere in S_1 consistent with f^2

at t_1 . We'll say that S_t imposes an ordering on logical space, " \leq_t ", which we can interpret as follows: for any $w, v \in \mathcal{I}^D$, $w \leq_t v$ just in case w is as close, or closer, of a possibility than v according to the deliberator's conscious information at t . Integrating a fragment of information into one's conscious information results in conscious information coming to consist of the \leq_t -closest information included in the fragment. To add some precision to the proposal, we can better define talk of "spheres" and "closeness". Following the work of Grove (1988), we will say that S_t is a system of spheres centered on i_t just in case S_t obeys the following conditions:

- S1:** For any $S, S' \in S_t$, $S \subseteq S'$ or $S' \subseteq S$
- S2:** $i_t \in S_t$ and for all $S \in S_t$, $i_t \subseteq S$ (i_t is the smallest sphere of S_t)
- S3:** $\mathcal{I}^D \in S_t$ (\mathcal{I}^D is the largest sphere of S_t)
- S4:** If $P \subseteq \mathcal{I}^D$ and $P \cap S \neq \emptyset$ for some $S \in S_t$, then there is some $S' \in S_t$ such that $P \cap S' \neq \emptyset$, and for all $S \in S_t$ where $P \cap S \neq \emptyset$, $S' \subseteq S$ (if a sphere intersects P , then there is a smallest sphere that intersects P)

The results of (Grove, 1988) also show that a system of spheres satisfies S1 – S4 just in case there is an ordering, \leq_t , that satisfies S1 \leq - S4 \leq .

- \leq_1 : For any $w, v \in \mathcal{I}^D$, $w \leq_t v$ or $v \leq_t w$ (connected)
- \leq_2 : For any $w, v, r \in \mathcal{I}^D$, if $w \leq_t v$ and $v \leq_t r$, then $w \leq_t r$ (transitive)
- \leq_3 : For any $w, v \in \mathcal{I}^D$, $w \in i_t$ just in case $w \leq_t v$ (all w 's in i_t are the \leq_t -closest points)
- \leq_4 : For any $P \subseteq \mathcal{I}^D$, there is some $X \subseteq P$ such that for all $w \in X$ and $v \in P$, $w \leq_t v$

This equivalency means that we can represent the conditional elements of conscious information either using a system of spheres or the corresponding ordering relation. We will opt for the later and say that a deliberator's conscious information consists of $\langle i_t, \leq_t \rangle$. The crucial change is that conscious information is no longer captured by i_t alone. Context permitting, we might sometimes use “conscious information” to refer to a deliberator's non-conditional conscious information i_t . But, we will only use “conditional conscious information” to refer to \leq_t .

One might worry that our new notion of conscious information no longer tracks that which we set out to track. After all, a deliberator need not reflect on her conditional conscious information. Consider a deliberator who consciously attends to *someone from the Mueller investigation has leaked* and *Putin is the Puppet Master*. After reflecting on her memory that there was no leak, her conscious information retains the information *Putin is the Puppet Master*. It's entirely possible that while integrating her memory into her conscious information, she never reflects on a conditional like *were there no leak, Putin would still be the Puppet Master*. In what sense, then, is it appropriate to treat such a conditional as part of her conscious information? The answer, I think, is to think of the conditional as part of the structure of her conscious information. It helps explain how she is poised to use and revise her conscious information.

Here, it might be illustrative to draw a distinction between conditional conscious information and conditional beliefs which serve no role in structuring conscious thought. Imagine the deliberator just discussed has the conditional belief *when tying boots, loop the laces like-so*. Her conditional belief helps explain and structure her non-reflective, shoe-tying behavior – we may imagine that when tying her boots, but not other footwear, she tends to loop the laces like-so. Let's assume that like her conditional conscious information *were there no leak, Putin would still be the Puppet Master*, she never reflects on her conditional belief. Neither is an object of reflection. Notice however that her conditional belief serves no role in shaping her conscious thought. It's not predictive of how she is likely to revise her conscious information in the face of incoming information. This of course is not true of her conditional conscious information. The lesson then is that while \leq_t might not be an object of reflection, it nonetheless serves an important role in structuring a deliberator's conscious information.

One advantage of understanding a deliberator's conscious information as i_t along with \leq_t is

that we now have a way of recognizing the sense in which two deliberators might both agree on how things currently are while disagreeing about certain counterfactual information. Consider for instance two deliberative processes, d and d' . At t , both deliberators consciously attend to the same non-conditional information: *someone from the Mueller investigation has leaked* and *Putin is the Puppet Master*. However, when integrating a memory that there in fact was no leak, only one retains her information about Putin. This difference can be captured by attributing to the deliberators conscious information which only differs in its conditional information: $i_t^d = i_t^{d'}$ and $\leq_t^d \neq \leq_t^{d'}$.

Given that we are now thinking of an agent's conscious information as $\langle i_t, \leq_t \rangle$, we have to reinterpret the integration operation. After all, we still want to understand a deliberator's conscious information at t as the result of updating her conscious information at $t - 1$ with a given information fragment. We can break our task into two parts: a) what should i_t look like given the update $\langle i_{t-1}, \leq_{t-1} \rangle \bullet i$? and b) what should \leq_t look like given $\langle i_{t-1}, \leq_{t-1} \rangle \bullet i$? We already encountered a proposal for (a). i_t will be the information in i which is \leq_{t-1} -closest to the deliberator's conscious information of c_{t-1} (i.e, it is that part of i which is in the smallest sphere representing the deliberator's conditional information). More formally, whenever $t \neq 0$, then

Constraint i_t : if $\langle i_{t-1}, \leq_{t-1} \rangle \bullet i$, then $i_t = \{w \in i : w \leq_{t-1} v, \forall v, v \in i\}$

Now let's look at (b), how to fix \leq_t given $\langle i_{t-1}, \leq_{t-1} \rangle \bullet i$. To do so we will draw from the results of (Darwiche and Pearl, 1997) for iterated belief revision. The general idea is that the constraints listed below inform how the conditional information of a deliberator's conscious information must change as a result of having integrated a selected fragment of information. Whenever $t \neq 0$, then

Constraint \leq_t : if $\langle i_{t-1}, \leq_{t-1} \rangle \bullet i$, then \leq_t satisfies $\leq_t 5 - \leq_t 8$:

$\leq_t 5$ For any $w, v \in i$, $w \leq_{t-1} v$ just in case $w \leq_t v$

$\leq_t 6$ For any $w, v \notin i$, $w \leq_{t-1} v$ just in case $w \leq_t v$

$\leq_t 7$ For any $w \in i$ and $v \notin i$, if $w <_{t-1} v$, then $w <_t v$

$\leq_t 8$ For any $w \in i$ and $v \notin i$, if $w \leq_{t-1} v$, then $w \leq_t v$

Constraints i_t and \leq_t let us define $\langle i_t, \leq_t \rangle$ in terms of $\langle i_{t-1}, \leq_{t-1} \rangle \bullet i$ resulting in the following reformulation of the Fragmented Implementation of QDAV.

Fragmented Implementation (second pass): Given $\langle \mathcal{I}^D, \mathcal{Q}, T, \bullet \rangle$, deliberative process d will consist of an information structure $\langle Q^d, \langle i_t, \leq_t \rangle, \mathcal{F} \rangle$ and the sequence d_0, d_1, \dots, d_n . $Q^d \in \mathcal{Q}$ is the deliberative state governing d , $\langle i_t, \leq_t \rangle$ is the deliberator's conscious information at t_0 , and $\mathcal{F}(Q^d) = \{f^1, f^2, \dots, f^n\}$ is the collection of information fragments accessible during every d_t . For each d_t :

1. if $t = 0$, then $d_t = \langle i_0, \leq_0 \rangle$
2. if $t \neq 0$, then for some selected fragment $i \in F$, $d_t = \langle i_{t-1}, \leq_{t-1} \rangle \bullet i$, where \bullet behaves according to Constraints i_t and \leq_t .

Part II

3.4. Selection operations

Now that we have a framework by which to understand the integration of some information fragment into conscious information, we can move on to explore the selection of an accessible fragment by which to integrate into one's conscious information. Here, our aim is to say something informative about which fragments of information the deliberator should consciously access and integrate into her conscious information –assuming that the integration will behave according to our integration operator \bullet .

One proposal which might strike one as worth considering is the idea that Selection Norms should march in step with \leq_t : a deliberator ought to select a fragment i just in case some part of i is as close, or closer, to i_t than any other fragment. Let's say that i is a \min_{\leq_t} -fragment just in case there is some $w \in i$, such that for any fragment $i^* \in F$ and $w' \in i^*$, $w \leq_t w'$. The proposed Selection Norm, then, is that a deliberator ought to select a \min_{\leq_t} -fragment to integrate into her conscious information. It's not hard to see why this won't always be a good deliberative strategy. Consider a deliberative process which begins as follows:

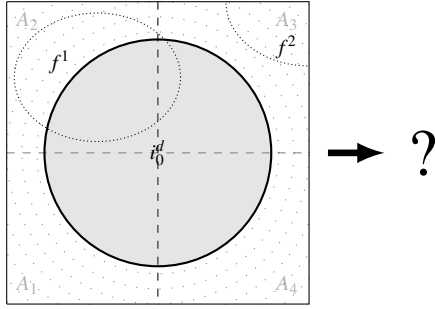


Figure 3.13: Which fragment to select first?

If we recommend that the deliberator select the \min_{\leq_t} -fragment for integration, then we recommend selecting f^1 . We know this because f^1 , but not f^2 , is consistent with i_0 . Of course, it would be a mistake to recommend against integrating with f^2 . Unlike f^1 , integrating with f^2 would ensure the resolution of the deliberative question at the next time slice of process.

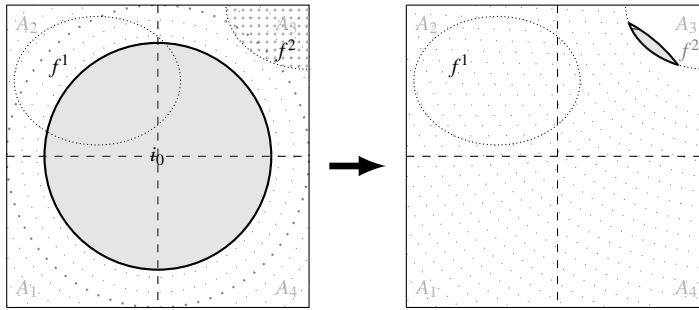


Figure 3.14

So we know then, that selecting the \min_{\leq_t} -fragment won't always ensure that one is proceeding in the most efficient fashion. Perhaps this result should not be surprising. Recall the Closeness Norm which says that it is better for a deliberative process to maximize the number of answers ruled out by the process. The adoption of some version of a Closeness Norm might seem to explain why integrating with f^2 is better than integrating with f^1 even though f^1 is the \min_{\leq_t} fragment. Consider this specification of the norm.

Closeness 1 All things considered, at d_t , select i for integration only if the result of (properly) integrating i into one's conscious information would bring her closer to resolving Q^d than integrating with any other fragment.

Closeness 1 seems plausible. Given that the whole point of deliberation is to resolve one's

deliberative question, it would only seem natural to adopt a norm that tells a deliberator to come as close as possible to resolving her deliberative question at each juncture of the deliberative process. But, upon further consideration, Closeness 1 is not an appropriate Selection Norm to adopt. To see why, let's see what happens when a deliberator obeys Closeness 1 when she occupies the following position.

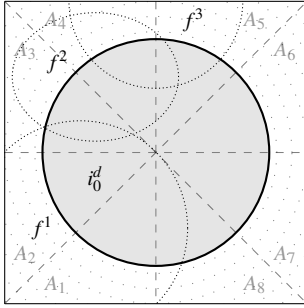


Figure 3.15: Closeness 1 dictates selecting f^3

If she follows Closeness 1, she will select f^3 for integration into her conscious information. This is because integrating with f^3 would rule out all but two answers to her deliberative question (A_4, A_5). By contrast, integrating with f^1 would leave open four answers to (A_1, A_2, A_3, A_8), and f^2 would leave open three answers (A_3, A_4, A_5).

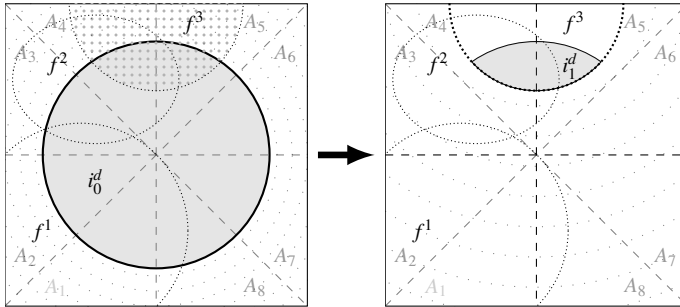


Figure 3.16: d during $t_0 - t_1$, obeys Closeness 1

What if she were again to obey Closeness 1? Doing so would require that she select f^2 for integration. This is because integrating with f^1 would actually expand the number of open answers to her deliberative question. At t_1 , the only open answers are A_4 and A_5 , but were she to integrate with f^1 , her conscious information would allow for A_1, A_2, A_3 and A_8 . This can be demonstrated by appealing to the constraints $\leq_t 1 - \leq_t 8$.

Let $w^1, w^2, w^3, w^8 \in f^1$ be contained in i_0 . We'll also stipulate that $w^1 \in A_1, w^2 \in A_2, w^3 \in$

A_3 , and $w^8 \in A_8$. $\leq_t 3$ dictates that for all $w \in i_0$ and for all w' , it must be the case that $w \leq_0 w'$. Thus,

1) For all w , it's the case that $w^1, w^2, w^3, w^8 \leq_0 w$

During $t_0 - t_1$, f^3 is integrated into $\langle i_0, \leq_0 \rangle$. $\leq_t 6$ requires that for any $w, w' \notin f^3$, $w \leq_0 w'$ just in case $w \leq_1 w'$. Because f^1 and f^3 are inconsistent, and $w^1, w^2, w^3, w^8 \in f^1$, we know that $w^1, w^2, w^3, w^8 \notin f^3$. From (1), it follows that for any $w' \notin f^3$ it must be the case that $w^1, w^2, w^3, w^8 \leq_0 w'$. Thus, in light of $\leq_t 6$,

2) For any $w \notin f^3$, it's the case that $w^1, w^2, w^3, w^8 \leq_1 w$.

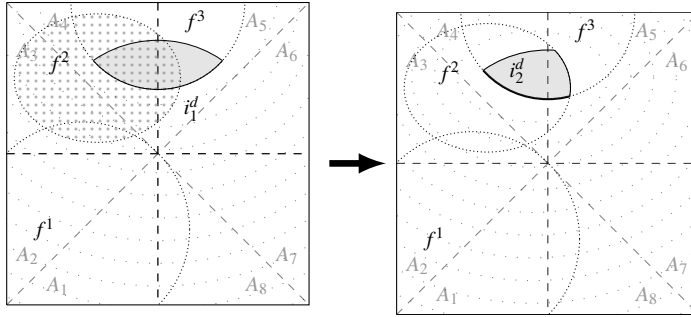
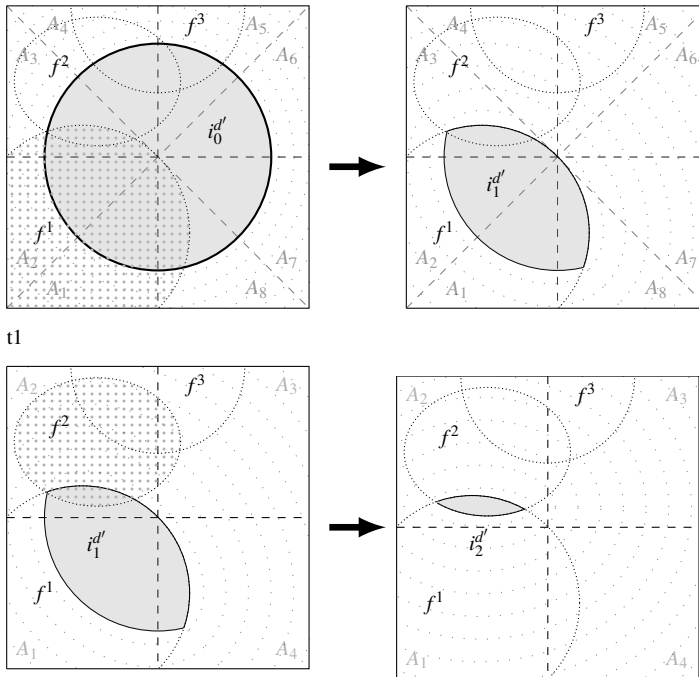
Now, suppose that from $t_1 - t_2$, the deliberator integrates f^1 : $\langle i_2^d, \leq_2 \rangle = \langle i_1^d, \leq_1 \rangle \bullet f^1$. By constraint i_t , $i_2^d = \{w \in f^1 : \text{for any } v \in f^1, w \leq_1 v\}$. Recall again that $w^1, w^2, w^3, w^8 \in f^1$. This means that $w^1, w^2, w^3, w^8 \in i_2^d$ so long as for any $w \in f^1$, $w^1, w^2, w^3, w^8 \leq_1 w$. Given that f^1 and f^3 are inconsistent, it follows from (2) that for all $w \in f^1$, $w^1, w^2, w^3, w^8 \leq_1 w$. Thus,

3) $w^1, w^2, w^3, w^8 \in i_2^d$

Remember that (3) describes the result of integrating f^1 during $t_1 - t_2$. It was stipulated that $w^1 \in A_1, w^2 \in A_2, w^3 \in A_3$, and $w^8 \in A_8$. At t_1 , the deliberator's conscious information only recognized two possible answers to her deliberative question, A_2 and A_5 . Thus, selecting f^1 to integrate during $t_1 - t_2$ would result in an expansion of the number of open answers. On the other hand, selecting f^2 to integrate during $t_1 - t_2$ would keep the number of open answers at 2. As such, Closeness 1 would require that the deliberator select f^2 and not f^1 to integrate during $t_1 - t_2$.

Closeness 1 leads to two open answers at t_2 . On the other hand, suppose the deliberator were to act in violation of Closeness 1 by first integrating with f^1 and then f^2 . Let's call that deliberation d' .

Clearly, this illustrates a case in which the most efficient strategy to resolve one's deliberative question violates Closeness 1. We offered Closeness 1 in an attempt to replicate on our model the Closeness Norm – the fairly intuitive idea that it is always rational to try to come as close as possible to resolving one's deliberative question. Assuming that the fundamental

Figure 3.17: d during $t_1 - t_2$, obeys Closeness 1Figure 3.18: d' during $t_0 - t_2$, violates Closeness 1

tenets of our model are correct, it would seem that the normative structure of deliberation is not so straightforward. Before abandoning the norm, it might be worth noting that Closeness 1 seems much more plausible when there are time constraints on deliberation. Consider the deliberator who occupies the position depicted in Figure 14. If it is unlikely that she will be able to perform more than one selection and integration operation, then obeying Closeness 1 will produce the better outcome. (Comparing d and d' during $t_0 - t_1$, we can see that the deliberation which follows Closeness 1 rules out the most answers.) Perhaps, then, the appropriateness of Closeness 1 (or other deliberative norms) will depend on the practical circumstances of one's deliberation.

Even if some version of a closeness norm will be apt for evaluating deliberation, such a norm will face the significant drawback of not being action-guiding. While in the grips of deliberation, one does not know which of her information fragments not yet attended to will bring her closest to resolving her deliberative question. It is only after she selects and then integrates a given fragment into her conscious information that she can observe its effects and determine how close it brings her to resolving her deliberative question. For this reason, whichever fragments she selects for integration cannot be explained by a norm that tells her to select the fragment which brings her closest to resolution. This is just to say that deliberators cannot be guided by a closeness norm.

Understanding why a closeness norm fails to be behavior guiding draws our attention to a possible modification to our framework. If we wish to secure selection norms which are action-guiding, our model will need to allow a deliberator to select a fragment of information for integration on the basis of some feature which she can have awareness of prior to selecting that fragment for integration into her conscious information. A plausible proposal to consider would be to “attach” to each fragment a label which the deliberator might access independently of her selecting and integrating the fragment. Consider for instance the proposal offered by (Bendana and Mandelbaum, *ming*) that, like mental files, belief fragments are associated with headings which are used to perform cognitive searches. Specifically, they envision headings as conceptualizations of environmental scenarios. If an agent enters a new cafe, Cafe Bar Bert, the information she learns at, and associates with, the cafe might be stored in a belief fragment with the heading BAR BERT. If she learns the meaning of “pusillanimous” during her visit, she will have an easier time recalling that it is a synonym for “cowardly” when primed with a description of the cafe.

For our purposes, it is not necessary whether the labels we attribute to belief fragments turn out to be environmental conceptualizations. It might very well be that for human deliberators, the labels attached to belief fragments are formatted as environmental conceptualizations while the cognitive architecture of a non-human deliberator employs fragment-labels with a different format. Of course, we also want to extend the idea of fragment-labels to whatever kinds of information fragments might play a role in deliberation (not just belief fragments). It might be that the labels for different kinds of information fragments are formatted differently. What

matters is that there are such labels and that they are formatted in such a way as to allow agents to scan the content of the header prior to the content of the fragment. This will allow us to make sense of selection norms which are action-guiding. For instance, a selection norm might instruct a deliberator deliberating about where to attend lunch to prioritize the selection of information fragments which have food-types or local restaurants in their heading.

3.5. Towards a virtue theoretic framework

An alternative strategy by which to investigate the normative structure of deliberation will follow a virtue theoretic approach.² Such approaches treat an agent's character traits as central to investigations of ethical and epistemic normativity. Similarly, we might explore the normative dimension of deliberation by focusing on an agent's "character" as a deliberator. Does she qualify as an efficient deliberator to the extent that she generally arrives at correct answers using minimal cognitive resources? Or does she often arrive at incorrect answers or require a multitude of update operations to resolve her deliberative questions? Consider though that there are potentially infinitely many deliberative questions one could deliberate over, some of which might be irrelevant to evaluations of deliberative efficiency. Furthermore, being efficient at resolving some questions might render one less efficient at resolving others. Thus, it looks like the character trait of deliberative efficiency will require some qualification.

When we talk about deliberative efficiency, I suggest we have in mind some privileged collection of deliberative questions by which to measure efficiency. The following analogy may help. Consider two libraries with identical collections. The first library organizes its collection into FICTION/NON-FICTION and various subcategories like MILITARY HISTORY. The second library organizes its collection according to publication year. While the first library will be more efficient at answering a question like, "Who was the greater general, Ghenghis Khan or Napoleon?", the second library will be more efficient at answering a question like, "How many books with 500 or more pages were written in 1985?" In the same way that the efficiency of each library depends on which questions we use as a measure, deliberative efficiency depends

²For virtue theoretic approaches to ethics see (Anscombe, 1958; Hursthouse, 1999), and for virtue theoretic approaches to epistemology see (Sosa, 2007; Zagzebski, 1996).

on which privileged deliberative questions will serve as measures for efficiency.

Perhaps the deliberative questions we privilege are those which a typical agent is likely to face over the course of her life. Perhaps they are those questions whose answers matter, in some sense or another, regardless of the likelihood of a typical agent deliberating over such questions. The questions to be privileged might be the product of evolutionary pressures, the sort of questions which one has been socialized to address, or some combination thereof. However we identify the collection of privileged deliberative questions, an efficient deliberator will be one whose information structures will render her efficient at resolving those very questions. Our strategy, then, will be to look at whatever information structures a deliberator has towards privileged deliberative questions and evaluate the efficiency of those structures.

Before proceeding, we will introduce a helpful simplification. Suppose we have identified $Q^1, Q^2 \dots Q^n$ as the privileged deliberative questions. Instead of checking piecemeal whether one's information structure for Q^1 is efficient, and whether one's information structure for Q^2 is efficient, and so on, it would be much more manageable if we only had a single privileged question to check. Fortunately, we can do this by taking the intersection of all the privileged deliberative questions. If we let $Q^{priv} = Q^1 \cap Q^2 \cap \dots \cap Q^n$, then we can think of Q^{priv} as that deliberative question which a deliberator will resolve just in case she resolves every privileged deliberation question.³ As an example, suppose that Q^1 and Q^2 are the privileged deliberative questions. In that case, we can visualize $Q^1 \cap Q^2 = Q^{priv}$ as Figure 18.

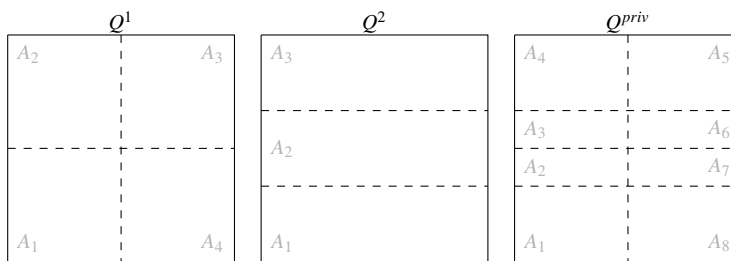


Figure 3.19: $Q^1 \cap Q^2 = Q^{priv}$

How do we use Q^{priv} to investigate deliberative efficiency? Recall that an information

³More technically, we can follow (Groenendijk and Stokhof, 1984) in defining the intersection of a set of questions $Q^1, Q^2 \dots Q^n$ as follows:

$$\bigcap_{i=1}^n Q^i = \{a_1 \cap a_2 \cap \dots \cap a_n \mid \text{for all } a_1 \in Q^1, a_2 \in Q^2, \dots, a_n \in Q^n \text{ where } a_1 \cap a_2 \cap \dots \cap a_n \neq \emptyset\}$$

structure has the form $\langle Q, \langle i_0, \leq_0 \rangle, \mathcal{F} \rangle$ and tells us about the relevant information which the deliberator has for deliberating about Q . Information structures by themselves don't resolve deliberative questions, but they do constrain one's possible deliberation about a question by fixing one's available information. So, when we talk about an information structure about Q^{priv} as being more or less efficient, we are really making claims about the the different deliberation made possible by that information structure. We will look at three senses in which an information structure might be considered efficient: how correct is the different deliberation made possible by a given information structure (correctness), how long does that deliberation take (length), and how much extraneous deliberation does the information structure make possible (volume).

3.5.1 Deliberative efficiency & correctness

Perhaps an obvious proposal to consider is that an efficient information structure ought to facilitate deliberation which correctly resolves Q^{priv} . Until now, we had put aside the issue of whether deliberation terminates at the “correct” answer because such considerations were largely irrelevant to constructing our descriptive framework of deliberation. And while discussing selection norms, we were mainly interested in norms which might serve a role in guiding deliberation. However, now that we are making evaluations from a virtue theoretic standpoint, it would not be inappropriate to think that a mark of deliberative efficiency is whether her information structure can supply a correct answer to a deliberative question.

Compare the two deliberative information structures STR^a and STR^b . Assume that A_1 is the correct answer to Q^{priv} .

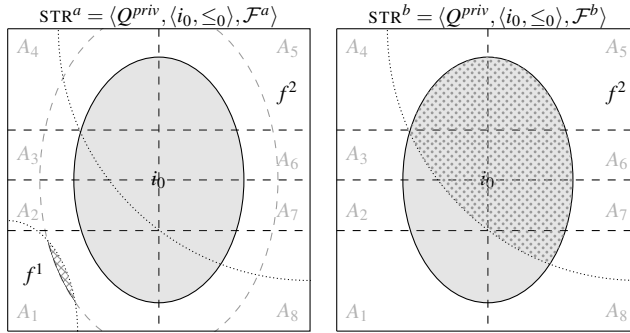


Figure 3.20: STR^a vs. STR^b

It should be fairly obvious that STR^a has the potential to correctly resolve Q^{priv} . STR^a allows for deliberation $d = \langle i_0, \leq_0 \rangle \bullet f^1$. Because the \leq_t -closest elements of f^1 are only consistent with A_1 (crosshatched region on the right), d arrives at the correct answer to Q^{priv} . On the other hand, the best that STR^b can provide is deliberation that arrives at an incorrect partial answer $(A_3 \vee A_4 \vee A_5 \vee A_6 \vee A_7)$ by way of $d' = \langle i_0, \leq_0 \rangle \bullet f^2$ (dotted region on the left). This is a clear sense in which STR^a is more efficient. It is important to keep in mind that in arriving at this determination, we are not claiming that a deliberator with information structure STR^a is guaranteed to correctly resolve Q^{priv} . After all, STR^a also allows for $d' = \langle i_0, \leq_0 \rangle \bullet f^2$. The point is just that a deliberator with F is capable of reaching the correct answer to Q^{priv} .

The proposal to understand deliberative efficiency in terms of correctness can be captured by assigning a “correctness score” to a deliberator’s information structure. To determine this score, we will look at all the deliberation made possible by the information structure and then pick-out that deliberation which most correctly resolves Q^{priv} . If that deliberation correctly resolves Q^{priv} , then we will assign a perfect score of 0 to that information structure. The less close that deliberation comes to correctly resolving Q^{priv} , the higher the score assigned to the information structure. Clearly, a perfect score of 0 should only be assigned to information structures like STR^a which make possible deliberation that rule out all, and only all, incorrect answers. By contrast information structure STR^b is deserving of the worse possible score as it only allows for deliberation which rules out the correct answer and no incorrect answers to Q^{priv} .

Even if a deliberator is unable to arrive at the correct answer to Q^{priv} she is still better off if her information structure allows her to arrive at a partial answer which is consistent with the correct answer. This is the sense in which information structures like STR^c and STR^d are less efficient than STR^a but more efficient than STR^b . Insofar as it is more efficient to occupy information structures which allow one to arrive as close as possible to the correct answer, STR^c has the advantage over STR^d (STR^c rules out more incorrect answers than STR^d).

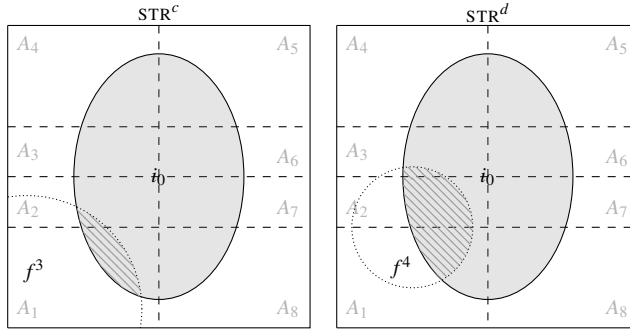


Figure 3.21: STR^c vs. STR^d

Let's take stock. We had said that the information structure STR^a is deserving of a perfect correctness score of 0. If the intuition pump just offered did its job, we should also be inclined to rank the four information structures just considered from most efficient to least efficient as follows: STR^a , STR^c , STR^d , STR^b . The correctness scores we assign to the four structures should reflect this ranking. Suppose we wanted a formula which assigns correctness scores to information structures in a manner consistent with these judgments. What should such a formula look like?

There are a host of different candidate formula we could consider. Which we ultimately settle on will require us stake out substantive positions. Do we care more about rewarding deliberative information for making possible deliberation that is consistent with the correct answer to Q^{priv} , inconsistent with incorrect answers, or strikes a balance? To illustrate this difference, we will consider three different formula, *Correctness Scoring* (CS) 1, 2, and 3. CS1 strikes a balance between rewarding information structures which make possible deliberation that is consistent with the correct answer while ruling out incorrect answers. CS2 favors information structures that make possible deliberation which is consistent with the correct answer, while CS3 favors ruling out incorrect answers.

To help formulate CS 1, 2, and 3, we will introduce the following terminology. For any possible deliberation d about Q^{priv} , $incr(d)$ will be the number of incorrect answers to Q^{priv} which are consistent with d 's resolution of Q^{priv} and $crct(d)$ will be the number of correct answers. Take for example, the only deliberation made possible by STR^c : $\langle \langle i_0, \leq_0 \rangle \bullet f^3 \rangle$. That deliberation is only consistent with the correct answer A_1 and one incorrect answer A_2 . Thus for that deliberation, $crct(d) = 1$ and $incr(d) = 1$.

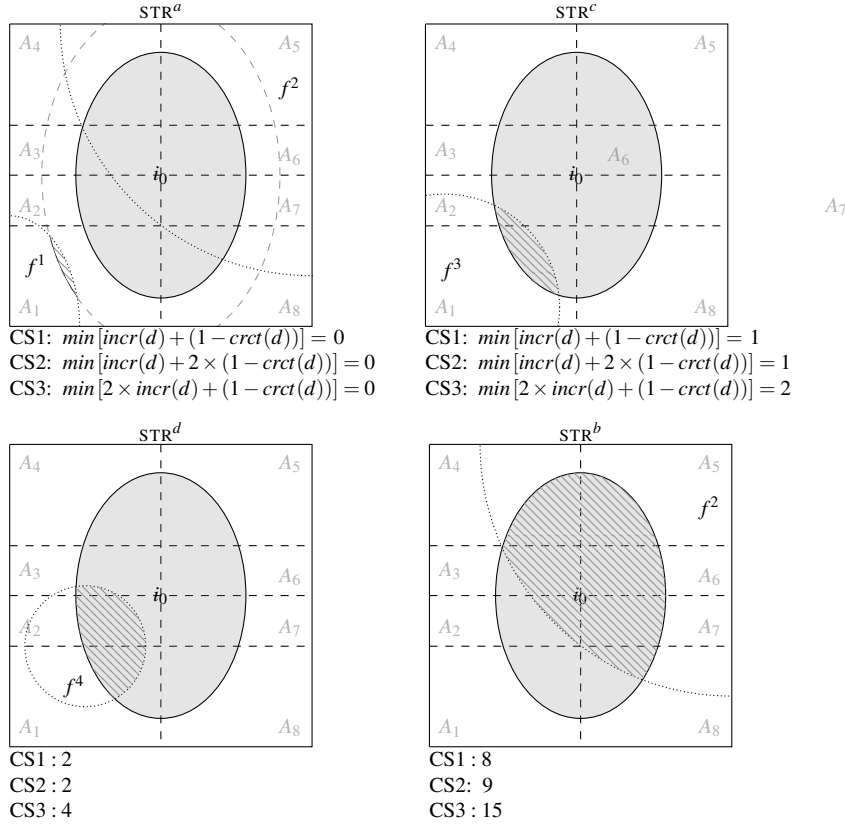
CS 1: The correctness score for a structure STR is the minimum value for $incr(d) + (1 - crct(d))$, where d is any deliberation made possible by STR

CS 2: The correctness score for a structure STR will be $\min[incr(d) + 2 \times (1 - crct(d))]$, where d is any deliberation made possible by STR

CS 3: The correctness score for a structure STR will be $\min[2 \times incr(d) + (1 - crct(d))]$, where d is any deliberation made possible by STR

Each of the three correctness scorings provide a formula for assigning a whole number to each deliberation made possible by an information structure. The smallest of those numbers is taken to represent the correctness score of the information structure.

The three scorings correctly assign 0 to STR^a . They also are consistent with our intuitions about how the four information structures considered should be ordered from most efficient to least efficient (in terms of correctness) as shown below.

Figure 3.22: STR^a vs. STR^b

Although CS1, 2, & 3 agree that the comparative ranking of the four information structures is STR^a, STR^c, STR^d, and STR^b, the difference between the three scoring formula comes out when we compare information structures like STR^d and STR^e.

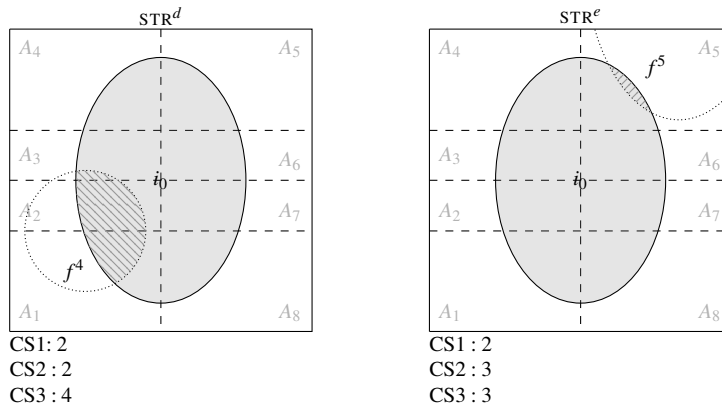


Figure 3.23: Comparing Closeness Scoring 1 - 3

With respect to helping a deliberator correctly resolve Q^{priv} , which information structure is more efficient? Correctness Scoring 1 is indifferent (both receive a score of 2). Correctness Scoring 2 says STR^d is more efficient than STR^e , and Correctness Scoring 3 says STR^e is more efficient. Which scoring formula better reflects deliberative efficiency would seem to depend on what we demand from different kinds of deliberation. Below, we will look at two cases, the first of which favors Correctness Scoring 2. The second case favors Correctness Scoring 3. Put another way, we will see that a deliberator in the first case would be better off with an information structure that looks like STR^d , while a deliberator in the second case would be better off with STR^e .

For the first case, suppose that a villain captures you and will only release you after you've chosen to drink from one of eight vials, seven of which contain a deadly poison. Not wanting to be imprisoned, you begin to deliberate about which vial to drink. Of course, knowing that only one vial does not contain the poison, your deliberation hinges upon properly identifying which vial contains the poison. Suppose that the correct answer to your deliberation is A_1 : *drink vial 1* (it's the safe vial). Would it be better for you to have an information structure that resembles STR^d or STR^e ? If you have a structure like STR^d , you will only partially resolve your deliberation by identifying five vials which you will not drink and that vial 1 is not among those five (alternatively, you determine that vial 1 is one-of-three vials which you might intend to drink). Such deliberation won't lead to your release, but it will secure your life and the possibility that future deliberation lets you arrive at the correct answer. If you have a structure like STR^e , your deliberation will result in you drinking one of the poisoned vials.

In the second case, imagine that the villain has already poisoned you prior to giving you the chance to earn your release by choosing one of eight vials to drink. One of the eight vials will have an antidote that is guaranteed to cure you. The other seven vials contain a less effective antidote which only offer a .5 chance of curing you. You again deliberate about which vial to drink where the correct answer of A_1 is to drink vial 1 (which will surely cure you). If you know that the poison in your system will soon take effect, you are best served by deliberation which results in an incorrect answer than deliberation which fails to resolve the question. After all, it would be better for you to drink one of the wrong vials and have a .5 chance of survival than drinking no vials and die. For this reason, an information structure like STR^d will be

less advantageous than STR^e . The deliberation made possible by STR^d only results in partial resolution, meaning that you fail to decide which vial to drink, resulting in your likely death. Far better for you to have a structure like STR^e which makes possible deliberation that would result in you choosing one of the vials that would give you a .5 chance of survival.

The two cases illustrate that the kind of scoring formula we prefer is likely to depend on the relative importance between avoiding incorrect answers and resolving one's deliberative question. My suggestion, then, is that different kinds of deliberation might push us to adopt different correctness scoring formulas.

3.5.2 Deliberative efficiency & length

Another sense in which some deliberative information structure might prove efficient is if it minimizes the length of the deliberation required to correctly resolve a question. By length, we have in mind the number of updates involved in a given deliberative process. For instance, although both STR^f and STR^g make possible deliberation which correctly resolve Q^d , STR^f has an advantage insofar as it makes possible deliberation which correctly resolves Q^{priv} using a single update. STR^g makes possible deliberation which resolves Q^{priv} using two updates.

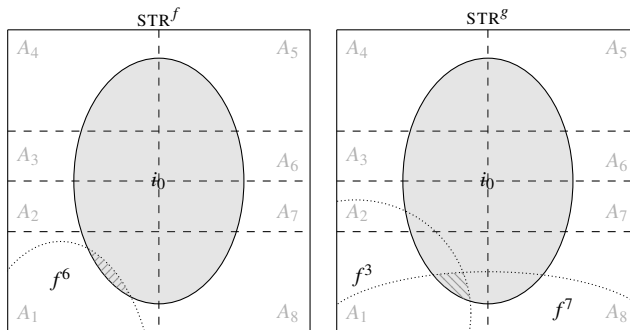


Figure 3.24: STR^f vs. STR^g

STR^f makes possible the deliberation $\langle \langle i_0, \leq_0 \rangle \bullet f^6 \rangle$ which arrives at the correct answer A_1 . To arrive at A_1 , a deliberator who occupies STR^f will either need to engage in the deliberation $\langle \langle i_0, \leq_0 \rangle \bullet f^3, \langle i_1, \leq_1 \rangle \bullet f^7 \rangle$ or the deliberation $\langle \langle i_0, \leq_0 \rangle \bullet f^7, \langle i_1, \leq_1 \rangle \bullet f^3 \rangle$.

To capture length as a measure of efficiency, we might try assigning a *correctness length score* to an information structure. This score will reflect the length of the most correct deliberation made possible by the information structure. A score of 0 indicates that the structure makes possible deliberation that correctly resolves Q^{priv} with the shortest length. Increasing scores indicate that the structure makes possible deliberation which fails to correctly resolve Q^{priv} with a short length. The shortest length a deliberation can have is 0 (the case in which the deliberator fails to select any fragments to integrate). But in order for deliberation to make progress towards resolving its deliberative state, it must rule out an answer by performing at least one update. Thus, the shortest length that productive deliberation might have is 1. By extension, a perfect correctness length score of 0 will be assigned to information structures which make possible deliberation with length 1 and which correctly resolve Q^{priv} . For instance, the structure STR^f in Figure 24 would be deserving of a perfect correctness length score of 0.

What formula should we use to assign a correctness length score? As with correctness scoring, there will be various candidates we might consider. I suggest we begin with whichever correctness scoring formula we find most suitable and then modify it to also reflect length. For, instance, suppose we find Correctness Scoring 1 suitable.

CS 1: The correctness score for a structure STR is the minimum value for $incr(d) + (1 - crct(d))$, where d is any deliberation made possible by STR

If we let $L(d)$ be the number of updates involved in deliberation d , we could transform CS1 by multiplying $incr(d) + (1 - crct(d))$ by $L(d)$. Call this Correctness Length Scoring (CSL) 1:

CLS 1: The correctness length score for STR is the $\min[L(d) \times (incr(d) + (1 - crct(d)))]$, where d is any deliberation made possible by STR

On the other hand, we might find that CSL1 weighs $L(d)$ too heavily. Perhaps a better option to consider is CSL 2 which adds, not multiplies, the length to $incr(d) + (1 - crct(d))$ by $L(d)$.⁴

⁴Why does CSL 2 add $(L(d) - 1)$ to $incr(d) + (1 - crct(d))$ instead of just adding $L(d)$? Remember that because we are not interested in deliberation with no updates, the shortest length any deliberation can have is 1. So, if we only added $L(d)$, the result would be a formula that assigns a correctness score length of 1 to possible deliberation which correctly resolves Q^{priv} in one update. To ensure CLS 2 assigns 0 to the most efficient deliberative processes we will add $(L(d) - 1)$

CLS 2: The correctness length score for STR is $\min[incr(d) + (1 - crct(d)) + (L(d) - 1)]$, where d is any deliberation made possible by STR

3.5.3 Deliberative efficiency as volume

Until now, we have been examining deliberative efficiency by honing in on one possible deliberation of a given information structure. In Section 2.1, we measured the efficiency of an information structure by looking at the most correct deliberation made possible by the structure, and in 2.3, we looked at the most correct and shortest deliberation made possible by a structure. But, another sense of deliberative efficiency might take into account how much deliberation made possible by an information structure is sub-optimal in the sense that it either fails to correctly resolve Q^{priv} or that it has a length greater than 1. To illustrate, compare STR^a and STR^f .

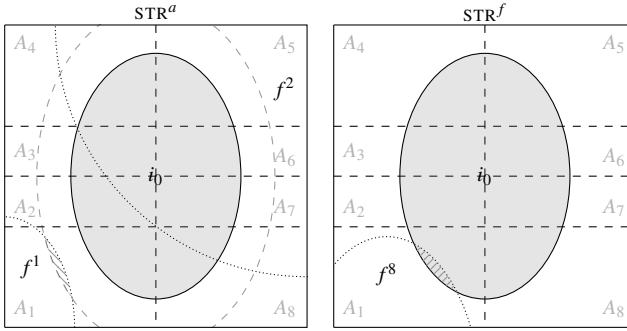


Figure 3.25: $STRC^a$ vs. $STRC^f$

Both STR^a and STR^f make possible deliberation that resolves Q^{priv} in the most optimal fashion. Specifically, STR^a makes possible the deliberation $\langle \langle i_0, \leq_0 \bullet f^1 \rangle \rangle$, while STR^f makes possible the deliberation $\langle \langle i_0, leq_0 \rangle \bullet f^8 \rangle$. Both deliberations correctly resolves Q^{priv} with the shortest possible length of 1.

Yet notice a crucial difference between the two structures – only STR^a makes possible deliberation which resolves Q^{priv} sub-optimally. STR^f only has one accessible information fragment, f^8 , so the only deliberation it makes possible is $\langle \langle i_0, \leq_0 \rangle \bullet f^8 \rangle$. But because STR^a has two fragments, f^1 and f^2 , it makes possible additional deliberation which fails to resolve Q^{priv} with length 1. If we list all the deliberation made possible by STR^a we have the following three

deliberative processes.⁵

$$\Rightarrow \langle i_0, \leq_0 \rangle \bullet f^1 \text{ (resolves } Q^{priv}, \text{ length 1)}$$

$$\Rightarrow \langle i_0, \leq_0 \rangle \bullet f^2 \text{ (only consistent w/ incorrect answers, length 1)}$$

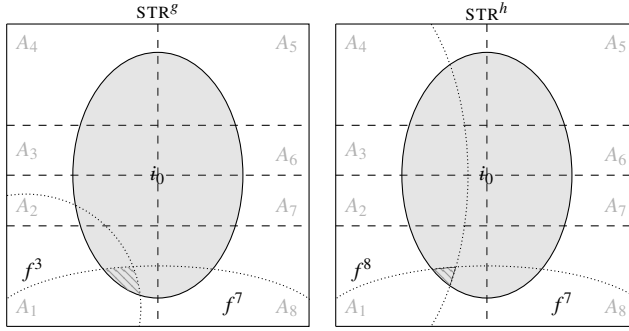
$$\Rightarrow \langle i_0, \leq_0 \rangle \bullet f^2, \langle i_1, \leq_1 \rangle \bullet f^1 \text{ (resolves } Q^{priv}, \text{ length 2)}$$

The deliberator who occupies the structure STR^a might deliberate in one-of-three ways, two of which fail to correctly resolve Q^{priv} with only a single update. In other words, two-out-of-three of STR^a 's possible deliberation can be thought of as a kind of excess. So long as it is the case that for each possible deliberation, the deliberator has greater than .0 chance of deliberating in that way, the two sub-optimal deliberations raise the likelihood that the deliberator will deliberate sub-optimally.

A natural thought to have when comparing STR^a with STR^f is that the disadvantage of STR^a is due to it having an information fragment, f^2 , which is inessential. Indeed, if f^2 were somehow to be “removed” from the structure, the deliberator would still be able to arrive at the correct answer to Q^{priv} using f^1 alone. Note, however that there are comparative judgments of deliberative efficiency we might draw in which the superiority of one information structure cannot be explained by an unnecessary information fragment. Consider structures STR^g and STR^h . Both have two fragments, and both fragments are essential for each structure to give rise to deliberation which correctly resolves Q^{priv} .

Both STR^g and STR^h make possible deliberation which correctly resolves Q^{priv} . The shortest length of that deliberation is 2 updates. Moreover, there is no fragment which is inessential. If either structure lacked one of its fragments, the structure would be unable to make possible deliberation that correctly resolves Q^{priv} . Nonetheless, given the total deliberation made possible by each structure, the deliberator who occupies STR^g as an advantage over the one who

⁵One might be tempted to conclude that the number of deliberative processes made possible by an information structure STR can be calculated in a straightforward way from the number of information fragments supplied by STR . Specifically, one might think that all the possible deliberation on STR is comprised by i_0 and all the different permutations of the fragments supplied by STR . So, if there are n number of fragments supplied by STR , then the total number of possible deliberation on STR should be $\sum_{k=1}^n \frac{n!}{(n-k)!}$. But the problem with this is that it fails to take into account that once a deliberator has arrived at an answer to her deliberative question, her deliberation is likely to end. Consider for instance that on STR^a , the above formula would include $\langle \langle i_0, \leq_0 \rangle \bullet f^1, \langle i_1, \leq_1 \rangle \bullet f^2 \rangle$ as possible deliberation. But, once a deliberator performs $\langle \langle i_0, \leq_0 \rangle \bullet f^1 \rangle$ she has already resolved her deliberative question and is unlikely to continue her deliberation.

Figure 3.26: STR^g vs. STR^h

occupies STR^h . This is because more of the deliberation made possible by STR^g arrives closer to a correct answer than the deliberation made possible by STR^h . To make this clear, let's list the possible deliberation of structures.

The deliberation made possible by STR^g :

$$\Rightarrow \langle i_0, \leq_0 \rangle \bullet f^3 \text{ (rules out 6 incorrect answers, length 1)}$$

$$\Rightarrow \langle i_0, \leq_0 \rangle \bullet f^7 \text{ (rules out 6 incorrect answers, length 1)}$$

$$\Rightarrow \langle i_0, \leq_0 \rangle \bullet f^3, i \bullet f^7 \text{ (resolves } Q^{priv} \text{ with length 2)}$$

$$\Rightarrow \langle i_0, \leq_0 \rangle \bullet f^7, i \bullet f^3 \text{ (resolves } Q^{priv} \text{ with length 2)}$$

The deliberation made possible by STR^h :

$$\Rightarrow \langle i_0, \leq_0 \rangle \bullet f^7 \text{ (rules out 6 incorrect answers, length 1)}$$

$$\Rightarrow \langle i_0, \leq_0 \rangle \bullet f^8 \text{ (rules out 4 incorrect answers, length 1)}$$

$$\Rightarrow \langle i_0, \leq_0 \rangle \bullet f^7, \langle i_1, \leq_1 \rangle \bullet f^8 \text{ (resolves } Q^{priv} \text{, length 2)}$$

$$\Rightarrow \langle i_0, \leq_0 \rangle \bullet f^8, \langle i_1, \leq_1 \rangle \bullet f^7 \text{ (resolves } Q^{priv} \text{, length 2)}$$

(Note that all of the deliberation made possible by each structure is consistent with the correct answer of A_1 .)

Two-out-of-four deliberative processes made possible by each structure correctly resolve Q^{priv} with a length of 2. While the other two deliberative processes made possible by STR^g

rule out 6 incorrect answers with a length 1, the other two made possible by STR^h do not do as well. One of the other possible deliberations on STR^h rules out 6 incorrect answers, while the other only rules out 4 answers. Quite plausibly, then, the deliberator who occupies STR^h will be at a disadvantage when compared to the deliberator who occupies STR^g .⁶ To capture this disadvantage, we will need a notion of deliberative efficiency which evaluates the correctness and length of total amount of possible deliberation on an information structure.

To measure the notion of deliberative efficiency at play here we might try proceeding as follows: assign an efficiency score to each deliberation made possible by an information structure and then calculate the average of those scores. To score a possible deliberation, we could borrow from our earlier discussions on calculating correctness length scores to an information structure. For instance, we suggested that one way of assigning such a score to an information structure was Correctness Length Scoring 1:

CLS 1: The correctness length score for STR is the $\min[L(d) \times (incr(d) + (1 - crct(d)))]$, where d is any deliberation made possible by STR

CLS 1 suggests that for any possible deliberation d , we can assign a correctness length score to d using the formula $[L(d) \times (incr(d) + 1 - crct(d))]$. This score will reflect how close d comes to correctly resolving Q^{priv} and how short its length is. Of course this is just one way of scoring a token deliberation. However we do, let's use $scr(d)$ for the efficiency score we assign to d . Once we have assigned a score to all the deliberation made possible by an information structure, the average of those scores will serve as the efficiency score for that structure. To put this more precisely, we'll let n be the number of different deliberative processes made possible by STR.

Volume Scoring: The volume efficiency score of STR is $\frac{1}{n} \times \sum_{i=1}^n scr(d_i)$, where n is the number of deliberations made possible by STR

⁶This point very much rests on the deliberator of each structure having roughly the same chance of instantiating each of the deliberative processes made possible by her structure. In other words, given the four deliberations made possible by each structure, each deliberation has approximately .25 chance of occurring. But, it might very well be that although each structure makes possible the four deliberations listed above, some have a much greater chance of occurring. If so, it might be that STR^g has no advantage over STR^h , or that STR^h actually has the advantage.

Chapter 4

A Question-Directed Attitude View for Epistemic Basing

4.1. Causal deviance & the epistemic basing relation

You see Spot run, and in so doing, come to believe that he has a shaggy brown coat. Your seeing Spot is said to be among the evidence or reasons for which you believe that Spot has shaggy brown coat. You are asked whether you enjoyed your dinner last night at the new restaurant, and you think back to last night and remember what you ate. Your belief that you ate and enjoyed the pad thai is based on your memory of having done so. The basing involved in such cases is epistemic: the relation which holds between a belief and the epistemic reasons or evidence for which the belief is held. In what follows, I will take *epistemic bases* to be mental states with propositional content – namely M is an epistemic base for S 's belief B iff for some evidence e , e is the evidence for which S believes B in virtue of S 's occupying M .¹

A very attractive kind of view analyzes basing in terms of causal relations. Such causal accounts generally maintain that M is an epistemic base for a belief B just in case M is at least a partial cause for the formation of B or M is a causal sustainer of B .² The idea is that your seeing Spot (and his coat looking brown to you) is an epistemic base for your belief because it plays a causal role in why you came to believe that his coat is brown. Your memory about the tofu pad thai is an epistemic base for your belief about it because your memory plays a causal role

¹It is appropriate to think of epistemic bases in terms of mental states with propositional content because it is generally held that in order for one's belief to be based on some evidence or reason, one must be cognitively related to the reason/evidence. After all, if it is possible for your seeing Spot to provide a basis for your belief that his coat is brown, it is either because your visual perception of Spot or his looking a certain way can serve as an epistemic base. If it's your visual perception, then we can treat a mental state as the basis for your belief. If it's Spot looking a certain way that serves as an epistemic base, it is only because his looking that way has been manifested in, or presented to you *via*, your visual perception of Spot. Either way, there should be a tight correspondence between a reason/evidence which is an epistemic base and the mental state which one must occupy in order for that reason/evidence to serve as an epistemic base.

²See (Swain, 1981) for an example of the former kind of causal account, and (Dretske, 1981) for an example of the latter. Also see (Korcz, 1997) for a broader discussion of the literature on epistemic basing.

in why you continue to believe, or are disposed to believe, that you ate the tofu pad thai. Not surprisingly, a causal analysis of basing must address the issue of causal deviance –a challenge familiar to causal accounts in various other domains.

As Davidson famously observed, causal accounts of intentional action must distinguish between deviant and non-deviant ways for an intention to cause an act, where only the former give rise to intentional action (1980). The problem, of course, is how to make sense of the distinction. It won't do to stipulate that an intention non-deviantly causes an act just in case the act qualifies as an intentional action, for that would be circuitous and *ad hoc*. Instead, the causal theorist of intentional action requires a principled means of demarcating between non-deviant and deviant ways for an intention to cause an act. Analogously, a causal account of the basing relation will require a principled means of demarcating between those mental states which cause/causally sustain a belief and qualify as epistemic bases and those mental states which cause/causally sustain a belief but are not epistemic bases.

To illustrate, suppose that your seeing Spot was caused by your desire to please Rhonda and your belief that by paying attention to her dog, you will gain favor with her. Both the belief and desire are part of the causal chain for why you formed the belief that Spot's coat is brown. But neither is an epistemic base for your belief that Spot's coat is brown. In consideration of the case involving recollection, suppose that through some odd quirk about you, your belief that you ate the tofu pad thai was causally sustained by your belief that you are a committed vegetarian. Perhaps if you stopped believing you were a committed vegetarian, you would suffer extreme physiological shock and disassociation that your mind would purge recent memories about what foods you consumed. If so, then your belief that you ate the tofu pad thai is causally sustained by your belief that you are a committed vegetarian. But surely, the later is not a basis for the former.³ What the two specifications of each case reveal is that the epistemic bases for a belief need not be co-extensional with the causes/causal sustainers of the belief. Consequently, a causal analysis of basing is confronted with its own problem of causal deviance.

A promising strategy has recently emerged that seeks to address the problem by appealing to manifestations of a disposition to believe, where a disposition to believe is a disposition

³See (Evans, 2013) for an analogous case

to form/sustain a belief.⁴ Label causal accounts of epistemic basing on which the bases of a belief are those mental states which cause/causally sustain a belief by manifesting a disposition to believe a *Causal-Dispositional Account* (CDA). To see how CDA hopes to address causal deviance, consider again your belief that Spot's coat is brown. Your belief was caused both by i) your seeing Spot and ii) your belief that by attending to Spot you will impress Rhonda, and your desire to impress her. Why is it that (i) but not (ii) is an epistemic base? Presumably one has a disposition to form beliefs about the color of objects from visual perceptions of those objects. One does not have a disposition to form a belief about the color of objects given a belief or desire about impressing someone. Thus, CDA can maintain that because only the causal relation between (i) and your belief that Spot's coat is brown manifests a disposition to believe, only (i) qualifies as an epistemic base for your belief. (Similar remarks hold *mutatis mutandis* for the tofu pad thai case.)

It is worth noting that any plausible version of CDA should deny that every cause of a belief which manifests a disposition to believe qualifies as an epistemic base. The qualification does work when we consider variations of the Spot case in which you are disposed to believe that Spot's coat is brown when you have a certain belief and desire about impressing Rhonda. Perhaps you know that Rhonda's love for Spot knows no bounds as she spends most of her waking day obsessively fixating upon him. If trying to gain favor with Rhonda will invariably lead one to visually attend to Spot, then it might be true of you that you have a disposition to believe that Spot's coat is brown when desiring to impress Rhonda. Or perhaps, only brown dogs are owned by overly enthusiastic dog-owners in your community. Under such conditions, you might have a disposition to believe that one's dog is brown whenever you believe that you can impress that person by visually attending to her dog. It looks like under these conditions, your belief that by looking at Spot you will impress Rhonda, and your desire to impress her,

⁴See for instance, John Turri's *Causal-Manifestation Account*, according to which mental state *M* is an epistemic base of belief *B* just in case *M*'s causing *B* manifests one of the believer's "cognitive traits". Here, we are to understand a cognitive trait as a "disposition or habit to form (or sustain) a doxastic attitude in certain circumstances" (2011, 17). More broadly, Sosa has suggested that in various domains (intentional action, perception, knowledge) in which a causal analysis is threatened by a species of causal deviance, the challenge can be met by understanding non-deviance in terms of those causes which manifest a competence (2015). Because for Sosa, a competence is a "disposition (ability) to succeed when one tries" (2015, 95), the epistemic bases of a belief will be identifiable with its causes which manifest a certain *epistemic competency* – namely a certain disposition to believe. Additionally, in (2018), Sylvan & Sosa defend a view of basing that analyzes the relation in terms of the manifestation of a competency to infer.

will be manifestations of a disposition to believe which cause your belief that Spot is brown. Nonetheless, it still seems wrong to say that your belief and desire about impressing Rhonda are epistemic bases for your belief about Spot. To make sense of this, CDA needs to distinguish between those manifestations of a disposition to believe which are epistemic bases and those manifestations which are not. Put differently, it needs to distinguish between non-deviant and deviant dispositions to believe, where only manifestations of the former serve as epistemic bases.

Turri and Sosa hope to secure a foothold by thinking about non-deviant dispositions to believe as “cognitive traits” or “epistemic competencies”, each of which reflects a stable capacity of an agent. More specifically, each thinks about the manifestation of a non-deviant disposition to believe as a kind of performance –an event which is evaluable because it is attributable to an agent or her cognitive system. You are criticizable if you see Spot in good lighting, under normal conditions, and you thereby come to believe that his coat is brown when the rest of us –were we in similar circumstances– would say that his coat is black. This is because forming a belief about the color of Spot’s coat in response to seeing Spot is something you do, or rather, something your cognitive system does. Similar remarks cannot be said of your believing that Spot is brown because of a belief/desire about impressing someone. Your believing that Spot is brown because you have a belief and a desire about impressing her is akin to your having a fly land on you because earlier you placed a banana in your bag. Neither is a performance on the part of your cognitive system.

CDA sought to address causal deviance by appealing to the notion of a cause that manifests a disposition to believe. But to prevent the problem of deviance from resurfacing at the level of dispositions to believe, the view now appeals to two kinds of epistemic performances. More specifically, CDA proposes that what is distinct about non-deviant dispositions to believe is that their manifestation are part of an epistemic performance to form or sustain a belief. Notice though that unless we have an independent grip on what an epistemic performance is, we are simply moving the lump under the rug from deviant causes/causal abstainers of belief to deviant dispositions to believe. Thus, the view relies on the following.

Epistemic Performance Principle: Were we to have a plausible, well-motivated account of

the epistemic performances of belief formation/sustainment, then distinguishing between the causes/causal sustainers of a belief which are part of such performances and those which are not would suffice as a non-ad hoc way of distinguishing between non-deviant and deviant causes/causal sustainers.

Unfortunately, versions of CDA are somewhat lacking when it comes to providing an account of what an epistemic performance might amount to. This paper aims to provide an account of epistemic performances on which the aim, or rather the function, of an epistemic performance is the settling of questions. To do so, the account analyzes epistemic performances as a kind of mental process – a set of mental states and the causal and functional relations between those states such that the process is attributable to a cognitive system. Specifically, epistemic performances are understood to be the mental processes by which a belief is formed or sustained.⁵

In the following section I outline a view of belief-forming processes on which the aim, or rather the function, of belief formation is the settling of questions. If belief formation is a cognitive endeavor to ascertain how the world is with respect to a question, then a process of belief formation will consist of a sequence of mental states, the content of which is relevant to some question. I then use the model of belief-forming processes to derive a view of belief-sustaining processes. As I argue in Section 4.3, supplementing CDA with this question-directed view of belief formation allows the causal theorist to distinguish between mental states which are causes/causal sustainers of a belief in virtue of being part of the process which forms/sustains the belief and mental states which are mere causes/causal sustainers of the belief. This should go a long way towards addressing causal deviance.

4.2. Belief-processes as question resolving

It is commonly acknowledged that one of the characteristic features of belief is that belief is truth-directed or that belief “aims” at truth (Williams, 1970; Wedgwood, 2002; Boghossian,

⁵It should be fairly clear why we are entitled to understand epistemic performances in terms of mental processes. An epistemic performance is a manifestation of a disposition to believe which is attributable to an agent or her cognitive system. Because a disposition to believe just is a disposition to form or sustain a belief, we can think about epistemic performances in terms of the mental processes by which a belief is formed or sustained.

2003; Shah, 2003; Sosa, 2007). How one further characterizes this aim is contentious. There is also some disagreement as to whether truth is in fact the aim of belief or whether there are not multiple aims of belief. I won't be committed to any particular story about how best to characterize the truth aim of belief or whether there are additional aims. But I do want to propose that belief aims at providing (true) answers to questions. It matters not whether one wants to read the proposal as an explication of the truth-aim of belief or as a separate aim.

What do I mean when I say that belief aims at answering questions? I do not mean that in order to believe p , a believer must have some question in the forefront of her mind. Instead, the idea is that questions play important explanatory roles in the epistemic performances by which beliefs are formed or sustained because the mental process whereby a belief is formed or sustained (a *belief processes*) aim at answering questions.

One way to motivate the proposal is to consider that processes of belief formation are topic sensitive: they are tasked with gathering, organizing, and formulating specific kinds of information. Generally, remembering what one ate yesterday tends to yield beliefs pertaining to topics like YESTERDAY and FOOD and not WWII BATTLES. Visual perception can be thought of as a process tasked with gathering information about surface reflectance properties and spatial relations of objects within one's visible field. Not surprisingly, visual perceptual beliefs tend to pertain to topics like COLOR, RELATIVE DISTANCE, and SIZE.

Questions provide a way of demarcating types of information and can thus help explain why a given type of belief-process tends to form certain kinds of belief. If the process of remembering what one ate yesterday aims to answer the question, "What did I eat yesterday?", then we should expect that remembering what one ate yesterday tends to form (or sustain) beliefs about food and the prior day. Similarly, the type of beliefs which tend to be formed on the basis of visual perception can be explained by characterizing visual perception as a process aimed at answering questions like, "What color are the objects in my visual field?", "What spatial relations do they bear to one another?" and "What size are they?". So the proposal is that belief-processes aim at answering questions, and types of belief forming processes can be individuated by the questions they aim to answer.

It is also worth noting that this aim of belief is not necessarily a person-level aim of the believer. When remembering what one ate yesterday, one may very well take oneself to be

trying to resolve a question about what one ate. But, when visually perceiving a scene, one need not take oneself as trying to answer some question about one's visual scene. So, when I say that believing aims at answering some question(s), I mean that characterizing belief-formation and sustainment as mental processes involves characterizing them as processes directed towards, regulated by, or tasked with the answering of a question. This is just to say that part of the functional role of a belief-forming process is the answering of a question (and that types of belief-forming/sustaining processes can be individuated by the questions they are tasked with answering). Let's call this aim of belief *Question Directedness*.

To the extent that mental processes are explained by their functional role, *Question-Directedness* should help explain elements of belief formation and sustainment. We already saw how thinking of belief-processes as question-directed might help explain why remembering what one ate yesterday and visually perceiving a scene tend to terminate in certain kinds of beliefs. But in order for *Question-Directedness* to play a role in more pointed explanations of believing, it should have something to say about why a given belief-process has the structure it does. A belief-process is a sequence of mental states culminating in a belief, along with the causal and functional relations driving the sequence of states. *Question-Directedness* should provide a story about a belief-process's underlying causal and functional structure. For *Question-Directedness* to play this role, we need a better grasp on what it could mean for a belief-process to have the functional role of answering a question. I suggest we think about belief-processes as involving a state that represents, or is in some way about, a question. Label a state that has a question as its content a *question-state*. We may now reformulate the functional role of belief forming processes as follows:

Question Directedness (QD):

For a given belief process p i) p involves a question-state, the content of which is a question Q , and ii) the function of p is the answering of Q .

Before it becomes clear how the proposal explains the structure of belief-processes –and in a way that will benefit CDA– some remarks on (i) and (ii). In Section 4.2.1, I will discuss (i) by outlining some of the distinguishing features of question-states. After which, we will be well-poised in 4.2.2 to examine how (ii) helps us understand the causal structure of a belief-process.

To streamline our discussion, I will focus on processes of belief-formation, until Section 4.3.2 where I extend the account to cover processes of belief-sustainment.

4.2.1 Questions as mental content

In many respects, recognition of questions as a kind of mental content is a natural extension of a well-established practice of treating questions as the semantic contents of an interrogative like, “Where can I find a newspaper?” Because such utterances do not seem to be in the business of communicating anything that could be said to be true or false, the content of such utterances is best thought of as something non-propositional. Following the insight of Jane Friedman (2013), we will do well to recognize a class of mental state which is the analog of interrogatives insofar as they have questions –not propositions– as content. The account of questions I will be adopting involves a standard, partition analysis according to which a question is an equivalence relation over logical space where the resulting partition comprises the set of complete answers to the question.⁶ In other words, a question carves up our space of possibilities into mutually exclusive and exhaustive regions corresponding to those propositions which are complete answers to the question. To illustrate:

Did you eat the tofu pad thai?

$\langle \text{I ate the tofu pad thai} \rangle, \langle \text{I did not eat the tofu pad thai} \rangle$

Did you eat the tofu pad thai, the red curry, or the vegetable dumplings?

$\langle \text{I ate none} \rangle, \langle \text{I only ate the tofu pad thai} \rangle, \langle \text{I only ate the tofu pad thai and the red curry} \rangle, \langle \text{I ate all three} \rangle, \langle \text{I only ate the red curry} \rangle, \langle \text{I only ate the red curry and the vegetable dumplings} \rangle, \langle \text{I only ate the vegetable dumplings} \rangle, \langle \text{I only ate the vegetable dumplings and the tofu pad thai} \rangle,$

What did you eat?

$\langle \text{I ate none} \rangle, \langle \text{I only ate } O_0 \rangle, \langle \text{I only ate } O_0 \text{ and } O_1 \rangle, \langle \text{I only ate } O_0, O_1, \text{ and } O_2 \rangle, \dots \langle \text{I only ate } O_0, O_1, O_2, \dots \text{ and } O_n \rangle, \langle \text{I only ate } O_1 \rangle, \langle \text{I only ate } O_1 \text{ and } O_2 \rangle, \dots$

⁶More formally, if \mathcal{W} is our construction of logical space, then for a given question Q , \mathcal{W}/Q is the set of equivalent classes $\{\mathcal{A}_0, \mathcal{A}_1, \dots, \mathcal{A}_n\}$ where every \mathcal{A} is a complete answer.

$O_2\rangle, \dots \langle I \text{ only ate } O_1, O_2, \dots \text{ and } O_n\rangle, \dots \langle I \text{ only ate } O_{n-1}\rangle, \langle I \text{ only ate } O_{n-1} \text{ and } O_n\rangle, \langle I \text{ only ate } O_n\rangle$ (For possible objects, $O_0, O_1, O_2 \dots O_n$)

Notice that so-called *partial answers* are not included in a question's answer set. Concerning (Q2), a proposition like $\langle I \text{ did not eat the red curry} \rangle$ is a partial answer because it is consistent with two or more complete answers but inconsistent with at least one complete answer. We will say that a body of information resolves a question just in case it entails a complete answer, and it partially resolves a question just in case it entails a partial answer (*i.e.*, it is inconsistent with at least one complete answer).

A question like (Q1), which only has two complete answers, is a binary question. Notice however that more complex questions, like (Q2) and (Q3), can be formed by stringing together various binary questions. It should be apparent that the answer set for (Q2) can be constructed from the answer sets for (Q1), $\langle \text{did you eat the red curry?} \rangle$ (Q1.2), and $\langle \text{did you eat the vegetable dumplings?} \rangle$ (Q1.3).⁷ Given the complexities of our belief-forming processes, many of the question-states that drive such processes will concern a question equivalent to some long, messy string of binary questions.⁸ For this reason, we must be careful to distinguish the question which is the content of a question-state from whatever interrogative we use to refer to the question-state. I refer to a question-state by way of an italicized interrogative. For instance, we might refer to the question-state involved in my belief-formation about what I had for dinner as *did I eat the tofu pad thai?*. But the content of that question-state probably will not be the question $\langle \text{did I eat the tofu pad thai?} \rangle$. Instead, the content will be a question will probably involve various presuppositions about the conditions under which I did or did not eat the tofu pad thai.

⁷More technically, we could follow Groenendijk & Stokhof (1984) in defining a question like (Q2) as the partition created by the intersection of the partitions of (Q1), (Q1.2), and (Q1.3)

⁸Take for instance, a visual perception which gives rise to a belief. Such a belief-forming process will involve a question-state which we may label *what are the objects within my visual field like?* Perhaps its content can be broken down into $\langle \text{what are the surface reflectance properties of the objects within my visual field?} \rangle$ and $\langle \text{what are the spatial relations among the objects within my visual field and myself?} \rangle$, and each of these questions can be broken down into a series of binary questions about each object within one's visual field. Of course specifying the particular bundle of polar questions which is the content of *what are the objects within my visual field like?* is beyond our reach. Knowing the answer set for that bundle of polar questions would be tantamount to having a comprehensive and correct account of the possible visual properties an object could have.

We can now gain a better understanding of question-states and their cognitive role by drawing some useful contrasts with the propositional attitudes. A propositional attitude involves a particular representation of the world; believing that I ate the tofu pad thai represents the world as such. Instead of representing a particular possibility, a question-state represents alternative possibilities of how the world might be, and in so doing, provides a cognitive system with a way of structuring information on a given subject matter or topic. The need for countenancing the propositional attitudes is already well-established, but it is not yet clear why an explanation of the mind will require positing question-states. Obviously one reason which I hope will emerge by the conclusion of Section 4.3 is that question-states are central to explaining belief-formation (and sustainment). To offer some independent motivation, I also suggest that such states can play an important role in the mental economies of creatures blessed with fairly intricate cognitive systems yet cursed with limited processing power.

As an illustration, consider a creature, *C*, with a cognitive system capable of collecting information *via* multiple perceptual modalities, and capable of storing, retrieving, and attending to various kinds of information. If *C* is out looking for food, it won't help if *C* indiscriminately collects and processes information about every environmental feature representable by her perceptual apparatuses. Nor will it help if she indiscriminately retrieves and attends to every bit of her stored information. Assuming *C* has finite cognitive resources, it will be impossible for her cognitive system to instantaneously process every bit of cognitively accessible information. If she fails to prioritize the processing of information which is relevant to food sources, there will be a good chance a food source would scurry away or be snatched up by a competitor by the time her system has processed the relevant information. It would be far more advantageous if her cognitive system had a means of raising the cognitive salience of information pertaining to the gathering of food.

Question-states provide one such means. At least this is true if part of the functional profile of a question state is to prioritize the processing of information which supports or rejects a complete answer to the question state's content. In virtue of occupying the question state *where is the food source?*, *C* is poised to attend to memories about, and form new beliefs and other mental states about the location of food sources. Another way to put the point is to note that because questions essentially emphasize certain distinctions between how things might be

while ignoring others, we can think about a question as a subject matter or information-type.⁹ Given question Q , a proposition pertains to the subject matter of Q just in case it rules out a complete answer to Q . Part of the functional profile of a question state will be to raise the cognitive salience of a subject matter:

Subject Matter Salience

Part of the functional role of a question-state is to make cognitively salient and prioritize the processing of information which at least partially resolves the question-state's content.

4.2.2 Function of belief formation

According to clause (ii) of (QD), the function of a belief-process is the answering of a question-state. We will say that a token belief-forming process answers its question-state just in case the information involved in the process resolves the question of that question-state. What is a belief-forming process's information? A belief-forming process is a sequence of causally and functionally related mental states terminating in a belief. So, the "information" of a belief-forming process just refers to the propositional contents of the states involved in the process. The picture that emerges of belief formation suggests that the beliefs formed through a process of belief-formation are a measure of how close the process has come to resolving its question-state. Whenever the process resolves the question by arriving at a complete answer, a corresponding belief is formed. By partially resolving its question-state, a process generates a belief towards a partial answer.

The current proposal calls for the following specification of *Question Directedness* concerning belief-forming processes:

Question Directedness (belief-formation)

(QD^{BF}) For token belief-forming process p , i) p involves a question-state, the content of which is a question, Q , and ii) the function of p is to resolve Q which occurs just in case the propositional contents of p 's constitutive mental states entail a complete answer to Q .

⁹For discussion on Lewisian subject matters see David Lewis (1988b; 1988a), and for discussion on information types, see Ivano Ciardelli (2018)

(QD^{BF}) allows us to explain processes of belief-formation in much the same way that the contours of a conversation between an inquirer and respondent can be explained by citing certain Gricean Maxims and attributing to the speakers the conversational goal of resolving the inquirer's question.¹⁰ Here, the inquirer's question is resolved, and the conversational goal is met, just in case an answer is entailed by the conversational common ground – that shared body of information accepted by the speakers for the purposes of the conversation. This analysis allows us to understand the various conversational moves between inquirer and respondent as a joint pursuit to communicate information, the effect of which is to update the common ground until it resolves the inquirer's question.

The most obvious way in which the common ground can be updated to resolve the inquirer's question is for the respondent to explicitly communicate a complete answer:

- 1) What did I eat for dinner last Friday?
- 2) You ate the tofu pad thai and nothing else.

Another way for a respondent to resolve a speaker's question is by communicating a proposition which is not a complete answer, but when added to the common ground entails a complete answer. Suppose an inquirer and respondent both know that the inquirer only eats vegetarian dishes. Such a respondent could then resolve the inquirer's question as follows:

- 2) What did I eat for dinner last Friday?
- 3) The tofu pad thai was the only vegetarian dish available last Friday.

Notice that even if a respondent is unable to lead the inquirer to a complete answer, she can still help the inquirer make headway by helping her to partially resolve a question. This occurs when the common ground is updated in such a way as to entail the negation of a complete answer while failing to entail a complete answer (*i.e.* a partial answer):

- 4) What did I eat for dinner last Friday?

¹⁰In what follows, I briefly sketch a view in the spirit of Craig Roberts' account (2012). While I cannot do justice to the richness of the view here, it should be noted that the account utilizes the notion of a "question under discussion" to provide a pragmatic analysis for all discourse and not just those conversations in which a speaker asks an explicit question.

5) I only remember that you didn't eat the curry.

Or, supposing again that both speaker and respondent know that the speaker only eats vegetarian dishes:

6) What did I eat for dinner last Friday?

7) I only remember that the curry was not a vegetarian dish.

In the exchanges above, updating the common ground with information provided by the respondent ensures that the common ground at least partially resolves the question under discussion. But, if the conversation reaches a point at which the respondent is unable to offer any helpful information by which to resolve the question, the inquirer might ask an additional question to signal that information pertaining to a certain subject matter might assist their conversational goal. In this way, an inquirer and respondent might resolve a question by stringing together utterances which, when considered in isolation, might be insufficient for the task.

8) What did I eat for dinner last Friday?

9) I only remember that you didn't eat the curry.

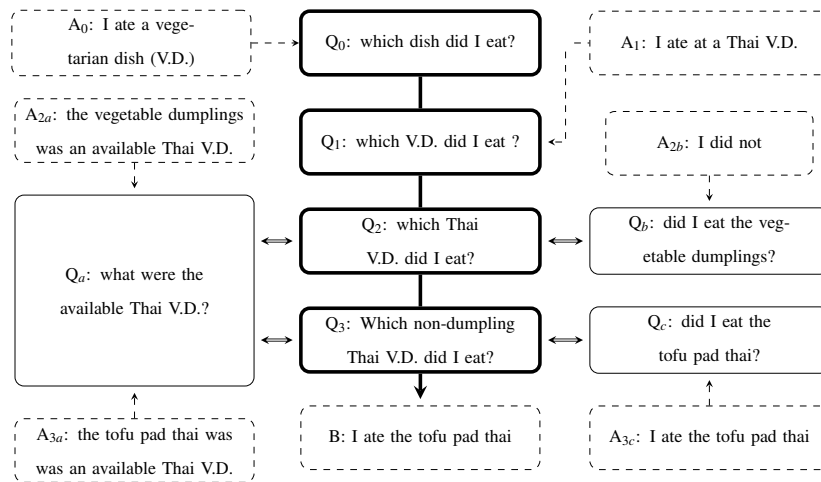
10) Well, do you remember which dishes were vegetarian?

11 The tofu pad thai was the only vegetarian dish.

12) Thanks! I'm a vegetarian, so I must have had that.

The general suggestion, then, is that a conversational exchange between inquirer and respondent can be understood as a series of utterances in the service of helping to resolve the inquirer's initial question. In much the same way, our Question-Directed account of belief-formation analyzes a process of belief formation as a sequence of mental states in the service of resolving the question of its main question-state. Following the main question-state, each state in the sequence will be an *auxiliary question state* –the content of which is a sub-question to help resolve the main-question state– or an *answer state* –a mental state with propositional content that resolves, or partially resolves, some prior question in the sequence. Each time the

contents of answer-states resolve, or partially resolve, the question-state, a corresponding belief is formed. To illustrate the proposal, the below diagram depicts a possible token belief forming process of remembering what one ate for dinner.



The function of the process is to resolve the main question-state *which dish did I eat?*, which as we have said will be a matter of resolving whatever question is the content of that state. The Q_0 -box depicts this question-state at the initial stage of the belief forming process. The series of subscripted Q -boxes along the bold downwards arrow represents different stages of the process. In this sense, each of these Q -boxes corresponds not to a distinct question-state but a transformation or update on the question-state driving the process. Notice that the bold downwards arrow (the “question-state arrow”) terminates at a “B-box”. This represents the belief that is the resolution of the question-state by the process. Because the content of B is a complete answer to Q_0 , the question-state of the process has been resolved, successfully fulfilling the question-directed function of belief-formation. Obviously, the process may or may not have satisfied a truth aim.

Among the mental states constituting a belief-forming process are *answer states*: clusters of attitudes with propositional content represented by “A-boxes”. Depending on the type of belief-forming process, a cluster could include beliefs, memories, perceptual states, and other relevant background assumptions and judgments. With respect to the process depicted in Figure 1, we can think of an A-box as a cluster of memories –either episodic or semantic– along with various assumptions concerning one’s diet, eating habits, local restaurants, *etc.* Notice that each A-box is a direct input to a Q -box, indicated by a dashed arrow traveling in the direction

of the Q -box. When an A -box is connected to a Q -box along the question-state arrow, a new Q -box is placed directly beneath it on the question-arrow, representing the effect of “updating” the main question-state with the information received from the answer-state. For instance, the effect of updating the question state at the Q_0 -stage of the process with the cluster of attitudes corresponding to A_0 : *I ate a vegetarian dish* is that it is no longer possible for the process to resolve its question-state with a complete answer according to which I ate a non-vegetarian dish. This is depicted by placing the box Q_1 : *which vegetarian dish did I eat?* beneath Q_0 : *which dish did I eat?*.

Why does every A -box resolve, or partially resolve, the question of a Q -box? Given that the function of a belief-forming process is to resolve its question-state, we should expect that at each stage of the process, the process will tend to select as informational inputs clusters of mental states that have contents which at least partially resolve the question-state. Notice also how the functional profile of a question-state facilitates the satisfaction of the process’s function. As per Subject-Matter Salience, a question-state raises the cognitive salience, and prioritizes the processing of, information which is relevant to resolving the question state. In other words, the question state of a belief-forming process makes cognitively salient the clusters of states corresponding to the A -boxes which the process draws on to try and resolve its question state.¹¹ (In this sense, we can think of the process of belief-formation as aimed at arriving at a complete answer to some question by collecting, sorting through, and manipulating information which is relevant to arriving at a complete answer.)

Finally, consider the branching Q -boxes, Q_a , Q_b , and Q_c . Unlike the Q -boxes along the question-state arrow, these boxes represent genuinely novel question-states and not just a transformation of the main question state *which dish did I eat?*. These *auxiliary question-states*

¹¹ Obviously, it is likely that a great many clusters will be included in the total body of information made salient by Q_0 : *which dish did I eat?*. I don’t have much to say about how such clusters are “formed” and made cognitively salient with the exception that the story will likely include some well-entrenched background beliefs and assumptions about the world. For instance, to say that A_0 : *I ate a vegetarian dish* represents a cognitively salient cluster to Q_0 might require that some of the candidate dishes which the question quantifies over are identifiable as non-vegetarian. I also won’t offer much of a story as to how, or why, one particular cognitively salient cluster is selected as an input to a Q -box at a given stage of the belief-forming process. I will say that as the process progresses and we move down along the question-state arrow, the amount of information made salient by the main question-state is narrowed. For instance, the salient information at the Q_0 -stage is more expansive than the information at the Q_1 -stage. For that reason, information which fully resolves the question-state is more likely to be selected at a latter stage.

could be deployed either because the belief-forming process has had difficulty locating cognitively salient information by which to resolve its main question-state (see Q_a) or because the auxiliary question-state facilitates a more efficient way of selecting cognitively salient information by which to resolve the question-state (see Q_b and Q_c).

Consider the process at the stage Q_2 : *which Thai vegetarian dish did I eat?* The lack of any connected A-box indicates that at that stage of the process, the process has not uncovered any cognitively salient information by which to even partially resolve its question-state. This temporary failure prompts the deployment of the auxiliary question-state Q_a : *what were the available Thai vegetarian dishes?*, raising the cognitive salience of information about which dishes were available for dinner on Friday (a subject matter not made salient by the main question-state). Once the process has partially resolved Q_a with A_{2a} : *the vegetable dumplings was an available Thai vegetarian dish*, the process feeds that information to its main question-state, indicated by the the double arrow connecting the Q_a and Q_2 boxes. The process can then use that information to pose another auxiliary question-state Q_b : *did I eat the vegetable dumplings?*.

Note that unlike Q_a , the auxiliary question-state of Q_b (and Q_c) does not make salient any information which was not already made salient by the main question-state. After all, every complete answer to the main question-state *which dish did I eat* entails an answer to Q_b : *did I eat the vegetable dumplings?* But because the converse does not hold, we can think of Q_b as targeting only some of the information made salient by the main question-state. Because the information made salient by Q_b is a more narrow body of information, resolving Q_b will be a much more manageable task. In this way, we can think of Q_b as the process deploying a sub-goal at a stage of the process in which resolving the main question-state has proven ineffective.

4.3. Question-directedness & basing

Let's return to causal accounts of the basing relation and the challenge presented by causal deviance. As noted in Section 4.1, a causal analysis of basing requires a non-ad hoc way of specifying how a mental state must cause, or causally sustain, a belief in order to qualify as an epistemic base for the belief. A causal dispositional account (CDA) offered the promising

strategy of identifying epistemic bases of a belief with those mental states which cause/causally sustain the belief by manifesting a disposition to believe. But because dispositions to believe are easy to come by, some causes/causal sustainers of a belief might manifest a disposition to believe without also qualifying as an epistemic base. For that reason, CDA restricted the epistemic bases of a belief to those causes/causal-sustainers which are part of an epistemic performance by which the belief is formed or sustained. So for CDA to succeed, it is in need of a principled means of distinguishing those manifestations of a disposition to believe that cause/causally sustain a belief *via* an epistemic performance from those manifestations which merely cause/causally sustain a belief.

4.3.1 Epistemic Performances of Belief-Formation

So, what exactly is an epistemic performance? At least with respect to those epistemic performances constituted by a belief-forming process, (QD^{BF}) generated the following answer: a sequence of causally and functionally related mental states consisting of a main question-state, possible auxiliary question-states, and answer-states which at least partially resolve the main question-state or auxiliary question-states. Because only answer states are mental states with propositional content, the view obeys the following constraint:

QD^{BF} Constraint:

For any state M with propositional content, M is a constituent of belief-forming process p only if M is part of an answer-state to p 's main question-state or an auxiliary question-state (*i.e.*, only if M at least partially resolves the main or auxiliary question-state)

Consider a version of CDA according to which a belief's epistemic bases are those causes which are part of the belief-forming process responsible for the belief. Given that an epistemic base is a mental state with propositional content, QD^{BF} Constraint tells us that only answer states may serve as epistemic bases. In other words, supplementing CDA with Question-Directedness about belief-formation will give us:

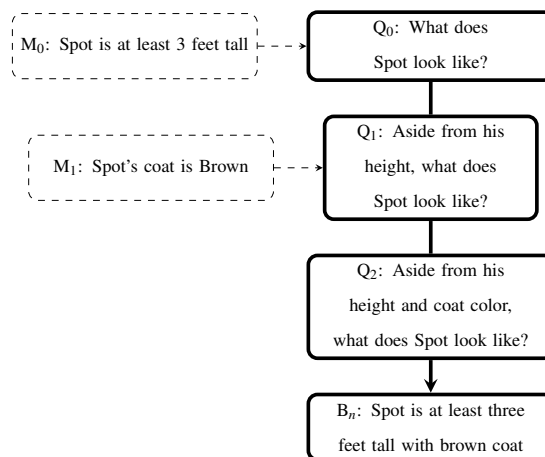
CDA + QD^{BF} :

Mental state M is an epistemic base for belief B just in case M causes B because M is part of an answer state of the belief-forming process responsible for B (*i.e.*, just in case M

causes B by at least partially resolving some question-state of the belief-forming process responsible for B).

To see how supplementing CDA with Question-Directedness helps, let us return to the case in which you believe that Spot is brown because you saw Spot, and you saw him because you both desired to please Rhonda and believed that you could do so by paying attention to Spot. Intuitively, your seeing Spot is an epistemic base for your belief that Spot is brown. Your belief or desire about what will impress Rhonda isn't. This is true, even if your circumstances are such that you are disposed to believe that Spot is brown whenever you desire to impress Rhonda and believe that you can do so by visually attending to Spot. According to CDA, this is because while your belief and desire about what will impress Rhonda might cause your belief about Spot by manifesting a disposition to believe, the manifestation is not part of an epistemic performance. Specifically, it is not part of the belief-forming process responsible for your belief about Spot. But your seeing Spot is part of such a process. Why is that?

According to $CDA + QD^{BF}$, an epistemic base for your belief that Spot is brown must be an answer-state of the belief-forming process responsible for your belief about Spot. We know from QD^{BF} that the belief-forming process responsible for your belief about Spot has the function of resolving a question state. Assuming you are not atypical, your belief-forming process will be driven by a question-state like *What does Spot look like?* It is clear how the contents of a visual perception of Spot and his brown coat are relevant for resolving the question state *what does Spot look like*:



A visual perception about the color of Spot's coat partially resolves the question-state which

your process of belief formation is driven to answer. By contrast, information about what would impress Rhonda and your wanting to impress her would clearly not serve to partially resolve the main question-state. It is also unclear what kind of auxiliary questions could be involved in a normal token process of belief-formation such that information about what impresses Rhonda would help resolve that question.¹²

4.3.2 Epistemic Performance of Belief-Sustainment

In order to supplement a version of CDA which recognizes causal-sustainers as epistemic bases, we have to extend our Question-Directed account of epistemic performances to cover performances of belief-sustainment. Like an epistemic performance of belief-formation, belief sustainment is also constituted by a mental process. There are, however, some notable differences between the two kinds of process. Unlike a process of belief-sustainment, the temporal parts of a process of belief-formation make a difference to the shape or structure of the process.¹³ Consider how the process whereby one forms the belief that Spot's coat is brown depends on the doxastic changes which occur during different temporal parts of the process. Only during the latter temporal part does the believer come to believe that Spot's coat is brown, and it is the doxastic changes which occur during preceding parts which explain the latter change. By contrast, the process whereby one sustains the belief that Spot's coat is brown does not correspond to doxastic changes in temporal parts.

We see, then, that a process of belief-sustainment is a continuous or persistent process that has the function of sustaining a prior belief. If so, then a plausible proposal for a question-directed aim of belief-sustainment can be given in terms of belief-formation.¹⁴

¹²Of course, it is not impossible to think of some auxiliary question-states for which such information might become relevant. Notice, however, that if we begin to add details to the case which explain why information about what will impress Rhonda answers some auxiliary question-state, which in turn helps answer the main question-state, we will begin to lose our grip on the intuition that a belief about what will impress Rhonda is not an epistemic base for the belief about Spot.

¹³This way of distinguishing between belief-forming and belief-sustaining processes maps unto Hellen Steward's useful distinction between mental events and processes (1997). A belief-forming process will correspond to a kind of mental event, while a belief-sustaining process will correspond to a mental process.

¹⁴One problematic feature of indexing processes of belief sustainment to processes of belief-formation concerns a possible tension with nativist theories of cognition according to which some information is innate. While not every nativist explanation must be spelled out in terms of innate beliefs, any that do will likely take issue with the proposed account of the epistemic performance of belief-sustainment. Consider for instance that one response to

Question Directedness of Belief-Sustainment (QD^{BS}):

For token belief-sustaining process p^* , p^* is defined relative to some belief-forming process p , where p has the function of resolving Q , and i) p^* involves a question-state, the content of which is Q , and ii) the function of p^* is to poise one's cognitive system to resolve Q in the same way p resolved Q .

Just like QD^{BF} gave rise to a picture of belief formation as a process consisting of question-states and answer states, QD^{BS} does the same for belief-sustainment. Not surprisingly, QD^{BS} will constrain the constituent states of a process of belief-sustainment in the following way.

QD^{BS} Constraint:

For any state M with propositional content, M is a constituent of belief-sustaining process p only if M is part of an answer-state to p 's main question-state or an auxiliary question-state.

Because epistemic bases must have propositional content, QD^{BS} Constraint tells us that the only constituent states of a belief-sustaining process which might be epistemic bases are answer states. So, if we consider a version of CDA which allows the causal sustainers of a belief to serve as epistemic bases so long as those causal sustainers are part of a belief-process, then Question-Directedness will provide for a clear way of identifying those causal sustainers which qualify as epistemic bases.

CDA + QD^{BS} :

Mental state M is an epistemic base for belief B just in case M causally sustains B because M is part of an answer state to the belief-sustaining process responsible for B .

4.4. Conclusion

To conclude, I want to raise one worry for Question-Directed views of belief formation and sustainment, and then note some extensions of the framework. First the worry: instead of solving

Noam Chomsky's Poverty of the Stimulus Argument (1965) is to maintain that language acquisition requires belief-like-representations of grammar (Stich, 1978). Fortunately, I suspect that QD^{BS} can be appropriately modified to accommodate various forms of nativism.

the problem of causal deviance, the view merely relocates the problem to the level of question-states. Take the belief that Spot is brown. One simple story for the formation/sustainment of that belief might involve a belief process that has the function of resolving a question-state with the content $\langle \text{Is Spot Brown?} \rangle$. But perhaps the belief is the product of a belief-process aimed at resolving $\langle \text{is Spot a dark shade of brown, a light shade of brown, or no shade of brown?} \rangle$. After all, on our Question-Directed framework, belief-processes yield beliefs by resolving, or partially resolving, a question-state. The indeterminacy encountered is especially acute given that for most questions Q , there will be a more fine-grained question Q' such that any complete answer to Q is a partial answer to Q' .

The worry, then, is that on the proposed framework, a given belief might be explainable by a range of belief processes with distinct question-states. Some of these belief processes might have potentially “deviant” question-states with content like $\langle \text{is Spot brown and what will impress Rhonda?} \rangle$. Notice that given some very plausible background beliefs, the belief that one will impress Rhonda by looking at Spot could serve as an answer-state to a belief-process aimed at resolving $\langle \text{is Spot brown and what will impress Rhonda?} \rangle$. This will frustrate a proponent of CDA from using Question-Directedness to explain why such a seemingly irrelevant belief will not qualify as an actual agent’s epistemic base for the belief that Spot is brown. Notice though that the issue we encounter is not the problem of causal deviance we began with. The issue is how to determine which question-states drive the belief-processes of actual agents.

Our difficulty in identifying which question-states are involved in the formation and sustainment of our beliefs seems to be another instance of the indeterminacy of mental content. Traditionally, interpretationist views of mental content have been restricted to propositional content (Dennett, 1988, 1981; Lewis, 1974; Davidson, 1973, 1974). But perhaps, a solution to the indeterminacy of question-states which mirrors interpretationist views of mental (propositional) content might help. The general idea will be to hold fixed the belief contents of a community of believers, determine which question-states are necessary for rationalizing the formation/sustainment of those beliefs, and then index “normal” types of belief-forming/sustaining processes to those question-states. Such a strategy might also offer a way to address the generality problem for process reliabilism – the problem of how to type a token belief process in order to determine its reliability. Indeed, it has been suggested (Comesaña, 2006) that a solution

to the generality problem can be procured by securing an account of the basing relation.

References

- Anscombe, G. (1958). Modern Moral Philosophy. *Philosophy*, Volume 33(124).
- Arpaly, N. and Schroeder, T. (2012). Deliberation and Acting for Reasons. *Philosophical Review*, 121(2).
- Bendana, J. and Mandelbaum, E. (forthcoming). The Fragmentation of Belief. In Kinderman, D., Onofri, A., and Borgoni, C., editors, *The Fragmented Mind*. Oxford University Press.
- Blackburn, S. (1984). *Spreading the Word: Groundings in the Philosophy of Language*. Clarendon Press.
- Block, N. (1990). Inverted earth. *Philosophical Perspectives*, 4.
- Boghossian, P. (2014). What is inference? *Philosophical Studies*, 169(1).
- Boghossian, P. A. (2003). The normativity of content. *Philosophical Issues*, 13(1).
- Bratman, M. (1987). *Intention, Plans, and Practical Reason*. Number 2. Center for the Study of Language and Information.
- Broome, J. (2009). The unity of reasoning. In Robertson, S., editor, *Spheres of Reason*. Oxford University Press.
- Broome, J. (2013). *Rationality Through Reasoning*. Wiley-Blackwell.
- Burge, T. (1996). Our entitlement to self-knowledge. *Proceedings of the Aristotelian Society*, 96(1).
- Camp, E. (2009). Two varieties of literary imagination: Metaphor, fiction, and thought experiments. *Midwest Studies in Philosophy*, 33(1).
- Camp, E. (2013). Slurring perspectives. *Analytic Philosophy*, 54(3).
- Camp, E. (2020). Imaginative frames for scientific inquiry: Metaphors, telling facts, and just-so stories. In Levy, A. and Godfrey-Smith, P., editors, *The Scientific Imagination*.
- Carballo, A. P. (2016). Structuring logical space. *Philosophy and Phenomenological Research*, 92(2).
- Cherniak, C. (1986). *Minimal Rationality*. MIT Press.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press.
- Ciardelli, I. (2018). Questions as information types. *Synthese*, 195(1).
- Ciardelli, I., Groenendijk, J., and Roelofsen, F. (2013). Inquisitive semantics: A new notion of meaning. *Language and Linguistics Compass*, 7(9).

- Comesaña, J. (2006). A well-founded solution to the generality problem. *Philosophical Studies*, 129(1).
- Darwiche, A. and Pearl, J. (1997). On the logic of iterated belief revision. *Artificial Intelligence*, Volume 89(1-2).
- Davidson, D. (1963). Actions, reasons, and causes. *Journal of Philosophy*, 60(23).
- Davidson, D. (1973). Radical interpretation. *Dialectica*, 27(3-4).
- Davidson, D. (1974). Belief and the basis of meaning. *Synthese*, 27.
- Davidson, D. (1978). Intending. *Philosophy of History and Action*, 11.
- Davidson, D. (1980). Essays on actions and events. In *Essays on Actions and Events*. Oxford University Press UK.
- Dennett, D. (1981). *Brainstorms*. The Harvester Press.
- Dennett, D. (1988). *The Intentional Stance*. MIT Press.
- Dretske, F. (1981). *Knowledge and the Flow of Information*. Oxford: Blackwell.
- Egan, A. (2008). Seeing and believing: perception, belief formation and the divided mind. *Philosophical Studies*, 140(1).
- Egan, A. (2018). De se pragmatics. *Philosophical Perspectives*, 32(1).
- Enoch, D. (2011). *Taking Morality Seriously: A Defense of Robust Realism*. Oxford University Press.
- Evans, I. (2013). The problem of the basing relation. *Synthese*, 190(14).
- Fodor, J. A. (1987). *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. MIT Press.
- Friedman, J. (2013). Question-directed attitudes. *Philosophical Perspectives*, 27(1).
- Friedman, J. (2019). Inquiry and belief. *Noûs*, 53(2).
- Gibbard, A. (2003). *Thinking How to Live*. Harvard University Press.
- Greco, D. (2014). Could KK Be OK? *Journal of Philosophy*, 111(4).
- Groenendijk, J. and Stokhof, M. (1984). *Studies on the Semantics of Questions and the Pragmatics of Answers*. PhD thesis, University of Amsterdam.
- Grove, A. (1988). Two Modellings for Theory Change. *Journal of Philosophical Logic*, Volume 17(2).
- Hamblin, C. L. (1958). Questions. *Australasian Journal of Philosophy*, 36(3).
- Harman, G. (1986). *Change in View*. Number 1. MIT Press.
- Hieronymi, P. (2013). The Use of Reasons in Thought (and the Use of Earmarks in Arguments). *Ethics*, 124(1).

- Hursthouse, R. (1999). *On Virtue Ethics*. Oxford University Press.
- Jackson, F. (1996). Mental causation. *Mind*, 105(419).
- Jackson, F. and Pettit, P. (1988). Functionalism and broad content. *Mind*, 97(July).
- Kolodny, N. (2005). Why be rational? *Mind*, 114(455).
- Korcz, K. A. (1997). Recent work on the basing relation. *American Philosophical Quarterly*, 34(2).
- Kornblith, H. (2012). *On Reflection*. Oxford University Press.
- Korsgaard, C. M. (2009). *Self-Constitution: Agency, Identity, and Integrity*. Oxford University Press.
- Lewis, D. (1972). Psychophysical and theoretical identifications. *Australasian Journal of Philosophy*, 50(3).
- Lewis, D. (1973). *Counterfactuals*. Harvard University Press.
- Lewis, D. (1974). Radical interpretation. *Synthese*, 27.
- Lewis, D. (1979). Attitudes de dicto and de se. *Philosophical Review*, 88(4).
- Lewis, D. (1982). Logic for equivocators. *Nous*, 16(3).
- Lewis, D. (1983). Postscripts to “radical interpretation”. In *Philosophical Papers*, volume 1. Oxford University Press.
- Lewis, D. (1988a). Relevant implication. *Theoria*, 54(3).
- Lewis, D. (1988b). Statements partly about observation. In *Philosophical Papers*, volume XVII, no. 1.
- Nagel, T. (1970). *The Possibility of Altruism*. Oxford Clarendon Press.
- Owens, D. (2008). Deliberation and the first person. In Hatzimoysis, A. E., editor, *Self-Knowledge*. Oxford University Press.
- Peacocke, C. (1998). Conscious attitudes, attention, and self-knowledge. In Wright, C., Smith, B., and Macdonald, C., editors, *Knowing Our Own Minds*. Oxford University Press.
- Roberts, C. (2012). Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics*, 5(6).
- Scanlon, T. M. (1998). *What We Owe to Each Other*. Harvard University Press.
- Shah, N. (2003). How truth governs belief. *Philosophical Review*, 112(4).
- Shah, N. (2008). How Action Governs Intention. *Philosophers’ Imprint*, 8.
- Shah, N. (2010). The Limits of Normative Detachment. *Proceedings of the Aristotelian Society*, 110(3pt3).

- Shoemaker, S. (1984). Some varieties of functionalism. In *Identity, Cause and Mind*. Cambridge University Press.
- Shoemaker, S. (1988). On knowing one's own mind. *Philosophical Perspectives*, 2(n/a).
- Smith, M. (1994). *The Moral Problem*. Blackwell.
- Sosa, E. (2007). *A Virtue Epistemology*, volume I. Oxford University Press.
- Sosa, E. (2015). *Judgment & Agency*. Oxford University Press UK.
- Stalnaker, R. (1984). *Inquiry*. Number 145. Cambridge University Press.
- Stalnaker, R. C. (1970). Pragmatics. *Synthese*, 22(1-2).
- Steward, H. (1997). *The Ontology of Mind: Events, Processes, and States*. Oxford University Press.
- Stich, S. P. (1978). Beliefs and subdoxastic states. *Philosophy of Science*, 45(December).
- Street, S. (2008). Constructivism about reasons. *Oxford Studies in Metaethics*, 3.
- Stroud, B. (2000). Practical reasoning. In Ullmann-Margalit, E., editor, *Reasoning Practically*. Oxford University Press.
- Swain, M. (1981). *Reasons and Knowledge*. Cornell University Press.
- Sylvan, K. and Sosa, E. (2018). The place of reasons in epistemology. In Star, D., editor, *Oxford Handbook of Reasons and Normativity*. Oxford University Press.
- Turri, J. (2011). Believing for a reason. *Erkenntnis*, 74(3).
- van Rooy, R. (2003). Questioning to resolve decision problems. *Linguistics and Philosophy*, 26(6).
- Velleman, D. (1989). *Practical Reflection*. Princeton University Press.
- Wedgwood, R. (2002). The aim of belief. *Philosophical Perspectives*, 36(16).
- Williams, B. (1970). Deciding to believe. In Williams, B., editor, *Problems of the Self*. Cambridge University Press.
- Williams, B. (1979). Internal and external reasons. In Harrison, R., editor, *Rational Action*, pages 101–113. Cambridge University Press.
- Yalcin, S. (2016). Belief as Question-Sensitive. *Philosophy and Phenomenological Research*, (1).
- Zagzebski, L. T. (1996). *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*. Cambridge University Press.