



University of Tennessee, Knoxville
**TRACE: Tennessee Research and Creative
Exchange**

Doctoral Dissertations

Graduate School

5-2019

Action as Essential Metaphysical Dependence

Jordan Timothy Baker
University of Tennessee

Follow this and additional works at: https://trace.tennessee.edu/utk_graddiss

Recommended Citation

Baker, Jordan Timothy, "Action as Essential Metaphysical Dependence. " PhD diss., University of Tennessee, 2019.
https://trace.tennessee.edu/utk_graddiss/5345

This Dissertation is brought to you for free and open access by the Graduate School at TRACE: Tennessee Research and Creative Exchange. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of TRACE: Tennessee Research and Creative Exchange. For more information, please contact trace@utk.edu.

To the Graduate Council:

I am submitting herewith a dissertation written by Jordan Timothy Baker entitled "Action as Essential Metaphysical Dependence." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Philosophy.

David Palmer, Major Professor

We have read this dissertation and recommend its acceptance:

Nora Berenstein, Eldon Coffman Jr., Bruce MacLennan

Accepted for the Council:

Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Action as Essential Metaphysical Dependence

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Jordan Timothy Ray Baker
May 2019

Copyright © 2019 by Jordan Timothy Ray Baker.
All rights reserved.

“Are my boots old? Is my coat torn?/
Am I no longer young, and still not half-perfect? Let me/
keep my mind on what matters,/
which is my work,/
which is mostly standing still and learning to be/
astonished.”

~Mary Oliver (2006)

This dissertation is dedicated to my first teachers—my parents, Tim and Sharron Baker—and my fellow student—my brother, Ethan Baker. Without your love, support, and wisdom I never would have made it this far.

ACKNOWLEDGEMENTS

Dissertations are strange. They are, in one sense, very narrow projects and so the list of people to thank who have directly contributed is short. Yet, they are also the culmination of an academic endeavor that can trace back over a decade. As such, acknowledging everyone who deserves it is difficult. Moreover, expressing the deep gratitude I feel towards those who have helped in this journey can only be faintly captured by words. That being said, an attempt must be made. To facilitate this, I choose to focus on those who have directly contributed to this dissertation's final form. To everyone else who I must, unfortunately, leave unnamed—thank you for everything, you deserve more than I can ever give.

First, and foremost, I thank my dissertation chair, David Palmer, for his tireless mentorship, good humor, and philosophical acumen. Working with David has made me a both better philosopher and writer. Moreover, in his teaching I was first confronted with the puzzles of human action in a way that was truly exciting. This dissertation is the fruit of those early classes and conversations. I also thank my committee members Nora Berenstain, E.J. Coffman, and Bruce MacLennan for their time and commitment. A special thanks to Nora whose teaching and mentorship early in my graduate studies fostered my desire to explore the intersection between metaphysics and philosophy science. Her guidance has fundamentally shaped my philosophical methodology, particularly my excitement for exploring the metaphysics of nature. In like manner, thanks to E.J. for his 2011 Survey of Epistemology course, which I took as an elective while finishing my musicology degree. I doubt I would have made the leap from studying music to studying philosophy without his encouragement and support.

I owe a debt of gratitude to all of the philosophy faculty at University of Tennessee, but some deserve special recognition for contributing to this project. I'd like to thank Jon Garthoff, Clerk Shaw, Josh Watson, Kristina Gehrman, and Mariam Thalos for their help. I would also like to thank retired professor John Hardwig, who has provided sage-like wisdom on all things regarding teaching and philosophy. Finally, I owe so much to our extremely competent administrative assistant Ashley Briggs and our former administrative assistant Susan Williams. Besides their immeasurable help with administrative, bureaucratic, and financial issues, their good humor and, frankly, their *kindness* have made graduate life more bearable than it had any right to be. Thank you both.

My research and writing were made easier thanks to generous support from the *Thomas Fellowship* and the *College of Arts and Sciences Dissertation Fellowship*. Thanks to all involved with that process.

I am lucky to have a graduate community that is supportive, cooperative, and philosophically adept. Each of them deserved to be thanked by name, but again I limit myself to those who contributed most directly to this project. To that end, I thank Michael Ball-Blakeley, Naomi Rinehold, Clayton Carden, Joseph Dartez, Michael Ebling, Samuel Webb, Tylor Cunningham, Jonathan Finnerty, Eddie Falls, Marlin Sommers, and Spencer Atkins. Your conversations and feedback have greatly improved this dissertation. Special recognition goes to former graduate student Devon Brickhouse-Bryson and my fellow cohort member Mary Helen Brickhouse-Bryson. I thank them for their philosophical input, but more generally for being

supportive friends and (in many ways) mentors as I shifted from studying music to studying philosophy. Also, a brief shout-out to former cohort members Mathew Reese and Albert Hu. I'm glad I was able to bounce some *very* early versions of these ideas off you both.

There are also a few non-philosophy individuals who have contributed to this research. I'd like to thank Sinead Doherty for her neuroscientific knowledge and practical counselors' insight regarding the relationship between mind and brain. It has been illuminating to learn from you and it helped me organize my own thoughts on the subject. I'd also like to thank my former student Jeremy Brunger for his relentless materialist reductionism, which served as a healthy check to my overly idealistic tendencies. You both (unknowingly) helped this dissertation be more carefully crafted.

Finally, a few people need special recognition:

My former housemates and good friends Robin and Lee Lovett Owen. You've helped me in more ways than I can count and conversations with both of you have informed the final shape of this project. Thank you both.

Thanks to Elizabeth Williams who (for some reason) decided to entangle her life with mine. Your philosophical rigor and loving care are *both* reasons that this dissertation reached completion. Thank you.

ABSTRACT

What makes an event *count* as an action? The standard answer to this question—causalism—claims that if an event is caused *in the right way* it counts as an action. Causal deviance objections, however, undermine the explanatory power of causalist accounts. Non-causal theories of action offer a promising alternative; however, they *also* raise a myriad of difficulties. Many non-causal arguments against causalism unintentionally lead to dialectical stalemates, which are methodologically undesirable and should be avoided whenever possible. I offer a theory between these two inadequate accounts that synthesizes the strengths of non-causalism with insights from agent-causal theories. I agree with traditional non-causalist that action explanations cannot be causally reduced; however, I also agree with causal theories that an *extrinsic* relation between the agent and the event makes the difference between mere events and actions. I call this account an “agency-first” theory of action since it neither reduces agency—as in causalism—nor does it ignore agency to focus on the intrinsic features of actions—as in non-causalism. Instead, I claim we must not lose sight of the *agent* when analyzing action and thus posit the non-causal, yet extrinsic, relation of essential metaphysical dependence to explain action *in terms* of agency without losing the distinctive character of either concept.

To this end, I claim that *essential metaphysical dependence* explains what makes an event count as an action by explaining how actions are grounded in agency. I first set-up the dialectic between causalists and non-causalists and raise objections to both views. I then describe the essential metaphysical dependence relation in detail and defend this account from several objections. Finally, the relation of dependence is commonly thought to be transitive, which entails a final significant objection—if actions depend on agency, and agency depends on non-agential forces, then actions are not *really* explained by dependence on agency. I argue, however, that plausible accounts of agency’s metaphysical emergence blocks the transitivity objection. I conclude that my agency-first theory adequately addresses what makes an event count as an action, while at the same time keeping the agent in view.

TABLE OF CONTENTS

Introduction.....	1
1. Dissertation Overview	1
2. Prior Commitments.....	3
3. A Final Note.....	7
Chapter 1: Three Non-Causal Accounts	8
1. Ginet.....	8
2. McCann.....	14
3. Goetz	31
4. Conclusion—An Important Similarity?.....	45
Chapter 2: Challenges for Non-causalism	47
1. Causalist Objections and Non-causalist Challenges	47
1.1 Causalist Objections: Randomness, Rational Explanation, and Origination.....	48
1.2 Non-causalist Challenges: Strong vs. Weak Non-causalism	70
2. Dialectical Stalemates and the Incomprehensibility Challenge.....	81
2.1 Dialectical Stalemates.....	84
2.2 The Incomprehensibility Challenge.....	90
Chapter 3: Action as Essential Metaphysical Dependence.....	96
1. A Brief Recap	97
2. Action as Essential Metaphysical Dependence.....	100
3. Dependence: An Overview	108
3.1 Existential Dependence.....	111
3.2 Metaphysical Dependence as Grounding.....	113
3.3 Essential Dependence	120
3.4 Dependence, Grounding, and Causation.....	124
4. Action as Essential Metaphysical Dependence-Redux.....	127
5. Action as Essential Metaphysical Dependence—Application.....	130
5.1 Satisfying Non-causal Desideratum.....	130
5.2 Addressing Causalist Concerns.....	134
6. Objections to Essential Metaphysical Dependence	141
6.1 Circularity Problem.....	142
6.2 Control Problem.....	148
7. Essential Metaphysical Dependence as Weak Non-Causalism	153
8. Conclusion	160
Chapter 4: Emergence and the Transitivity Objection.....	162
1. Emergence—History and Conceptual Development	163
2. The Collapse Problem and a Hybrid Account of Emergence	177
3. The Transitivity Objection and Emergent Agency	199
4. Conclusion—A Question of Plausibility?.....	211
Conclusion	220
1. Free Action and Essential Metaphysical Dependence	221
2. Animal Action and Essential Metaphysical Dependence	223
List of References	226
Vita.....	241

INTRODUCTION

What makes an event *count* as an action? What makes the difference between my lifting my arm and a gust of wind moving it instead? The standard answer to this question—causalism—claims that if an event is caused *in the right way* it counts as an action. Causal deviance objections, however, undermine the explanatory power of causalist accounts both in explaining what makes an event count as an action as well as explaining its unique connection to agency. Non-causal theories of action offer a promising alternative; however, they *also* raise a myriad of difficulties. Many non-causal arguments against causalism unintentionally lead to dialectical stalemates, which are methodologically undesirable and should be avoided whenever possible.

I offer a theory between these two inadequate accounts that synthesizes the strengths of non-causalism with insights from agent-causal theories. I agree with traditional non-causalist that action explanations cannot be causally reduced; however, unlike traditional non-causalists I agree with causal theories that an *extrinsic* relation between the agent and the event makes the difference between mere events and actions. I call this account an “agency-first” theory of action since it neither reduces agency—as in causalism—nor does it ignore agency to focus on the intrinsic features of actions—as in non-causalism. Instead, I claim we must not lose sight of the *agent* when analyzing action and thus posit the non-causal, yet extrinsic, relation of essential metaphysical dependence to explain action *in terms* of agency without losing the distinctive character of either concept.

1. Dissertation Overview

My argument is, roughly divided in two parts. The first half—articulated in chapters 1 and 2—sets-up the debate by first presenting three important non-causal views and considering their objections to causalism.

Chapter 1 focuses on three non-causal theories that together exemplify the standard non-causal views in the literature. First, the “weak” non-causalism of Ginet (1990), which allows for the possibility that our actions are caused. Second, the “strong” non-causalism of McCann (1998) and Goetz (2008), who both claim that our actions *must* be uncaused to be actions at all. In this chapter, I present these three views with minimal critical apparatus. The goal, for the sake of charity, is to present each view in a manner that encapsulates the author’s original presentation. If I believe an author’s argument is particularly opaque I may expand their ideas some, however, this is not the primary goal of chapter 1.

In contrast, chapter 2 explicitly considers causalist objections to these non-causal views. In this chapter, I emphasize and support the criticisms that the non-causalists raise against standard causal account while also defending them against common causalist objections. I conclude that while their objections *to* causalism are successful, non-causalist are vulnerable to several counter-arguments. I also raise a concern that non-causal arguments often lead to unproductive dialectical stalemates within philosophy of action.

The second half of my argument—found in chapters 3 and 4—articulates my positive account. Given the failures of both standard causal views and traditional non-causal alternatives, a successful account of action must synthesize elements from both theories.

To this end, in chapter 3 I articulate an “agency-first” account in which a non-causal, yet extrinsic, relation of *essential metaphysical dependence* explains what makes an event count as an action by explaining how actions are grounded in agency. I begin by recapping the conclusions reached through chapters 1 and 2, this leads to my initial definition of *essential metaphysical dependence*. I then spend much of chapter 3’s middle section attempting to clarify various terms used in this rough definition (e.g. dependence, essential, agential properties, etc.).

Once these terms are clarified I describe this relation in detail and apply it to the domain of action. I conclude by responding to several objections to my account.

Finally, in chapter 4, I note that the relation of dependence is commonly thought to be transitive, which entails a final significant objection—if actions depend on agency, and agency depends on non-agential forces, then actions are not *really* explained by dependence on agency. I argue that this “transitivity objection” is also unsuccessful. I argue that plausible accounts of agency’s metaphysical emergence provide an ontological “gap” that blocks the transitivity objection. I argue for this by first clarifying the concept of emergence. In particular, I engage with debates regarding whether emergence is a *unitary* concept that is primarily *epistemic* in scope or a *pluralistic* concept that is *metaphysical* in scope. I, ultimately, advocate for a hybrid view. With this account of emergence in hand, I apply it to the transitivity objection, showing that both strong and weak emergence provide enough of an ontological gap to allow my account to avoid the problem. I conclude with a brief section that bolsters the plausibility of my account by canvassing the various scientific domains wherein emergence is seen as a live option. Thus, my agency-first theory adequately addresses what makes an event count as an action, while at the same time keeping the agent in view.

2. Prior Commitments

As with any piece of philosophical writing, this dissertation depends on many background assumptions. These assumptions are taken for granted not because arguments surrounding them are settled, but merely because there is never enough time or space to say everything. I believe that where such assumptions directly motivate an argument, philosophers have an obligation to be transparent about what unargued claims they are utilizing. In this section, I do just that. I

consider three background assumptions that, though unargued for, inform and motivate this project.

(i.) Philosophical Dialectic and its Purpose

In chapter 2, I claim that the idealized goal of philosophical debate is to arrive at an intersubjective, stable, and reflective judgment about a topic in the hope that this judgment by its very intersubjectivity, stability and reflectiveness is most likely to be true. I further clarify this as arriving at truth via reflective consensus building among the various parties of a philosophical dispute. This claim about the *ideal* of philosophical dialectic is important, because it motivates a criticism of the current dialectic between causalists and non-causalists as being in a stalemate.

I was inspired to understand philosophical dialectic in this manner from a variety of sources: Lloyd's (1995) work regarding the relationship between intersubjectivity and objectivity in science; Rawls' (1951; 1971, pp. 19-22, p. 580; 2001, pp. 1-5) reflective equilibrium methodology as well as his claim that political philosophy can be understood as a kind of "reconciliation;" and Brewer's (2009) discussion of the value of dialectical activities, just to name a few. It is worth noting here that I'm not sure this is the *only* appropriate philosophical methodology. Indeed, I explicitly claim above that this is an ideal and as such certain non-ideal circumstances might require different methods. Furthermore, there might be debates that involve philosophy that are not themselves "philosophical debates" in the technical sense and thus require different methods. I hope, such a conception of philosophical dialectic, is plausible on its face.

(ii.) Metaphysics as Reconciliation

Sellars' (1962) famous distinction between the "scientific image" and "manifest image" runs throughout the background of my work. Metaphysics, as I practice it, is aimed at

“reconciling” the manifest image, our pre-theoretic understanding of ourselves, with the scientific image, the post-theoretic understanding of ourselves. Furthermore, a genuine reconciliatory project should, I believe, aim to maintain the unique features of each “image,” insofar as this is possible. Mere reduction is not reconciliation. In like manner, establishing the coherence and consistency of *one* image without ever considering (even minimally) how it fits with the other image is not reconciliation. That is, mere parallelism is not reconciliation.

This provides motivation for accepting the plausibility of certain views. All other things equal, if an account allows for *reconciliation* rather than *reduction* or *parallelism* then it is to be preferred. In my dissertation this assumption motivates my “agency-first” account. Of course, this principle is defeasible and I hope in practice such a method will not diverge greatly from other methods in metaphysics. But, I suspect taking the manifest image as something to preserve and reconcile rather than reduce may affect my judgments regarding the *plausibility* or *undesirability* of certain conclusions. I could see a critic claiming that I have an overly “pragmatic” view of justification, which is inappropriate for metaphysics in general and especially a metaphysical project that is attempting to be (in some sense) “naturalistic.” I have no extended response to this, I merely note that in some ways I am willing to bite this bullet—by my lights, *all* philosophy is in the business of discovering to what extent we can justifiably *humanize* the world (that is, reconcile ourselves to each other, to non-human animals, to nature itself, etc.) and thus some amount of practical or pragmatic justification is appropriate.

(iii.) *Explanatory Realism and Pluralism*

Questions regarding *explanation* loom large throughout my dissertation. Yet, I offer no explicit account of the nature of explanation, scientific or otherwise. This is simply because such a topic would be a dissertation unto itself. Given my already extensive explanation of

“dependence” and “grounding” to try and tackle the nature of explanation in even a cursory manner would be ill advised. As such, I will merely note that I take both a moderately *realist* and *pluralist* stance towards explanation.

Explanatory realism is, roughly, the claim that explanations give information about what determines the explanandum. More generally, it is the view that our best explanations tell us something about entities in the world and *this is why* they are successful as explanations. I take such a view to be relatively popular in metaphysics. For example, I believe form of explanatory realism to undergird so-called “indispensability” arguments for mathematical realism (e.g. Colyvan, 2001) and “no miracles” arguments for scientific realism (e.g. Lyons, 2003). The claims I make in chapter 3 regarding how metaphysical explanation is “backed” by some real relation depends on this assumption. It also plays a part in chapter 4, where I claim that one can infer facts about the world from failures of explanation. There are, of course, reasons to doubt this understanding of explanation and Taylor (2018) offers a good assessment of the state of the debate and a rigorous argument *against* such realism.¹

Explanatory *pluralism* is, roughly, the claim that there is more than one *type* or *kind* of explanation. This claim may seem obviously true, but it is a slightly more contentious claim when one is discussing explanations within the domain of science or the “natural” more broadly construed. Traditionally, *causal* explanations have often been taken to be exclusively salient for the domain of the natural.² There is an *ample* literature on this in both philosophy of science and metaphysics and it seems, happily for me, that the consensus is moving towards explanatory

¹ Though, importantly, Taylor (2018) defines explanatory realism as the view that *all* explanations give information about whatever metaphysically determines the explanandum (p 198). I would not want to hold that strong of a position, hence my *moderate* explanatory realism.

² I take the, so called, “Eleatic principle”—the principle that only things that are causally salient exist—to provide some of the background motivation here. For a critical assessment of this principle see Colyvan (1998).

pluralism (For example, Thalos. 2002; and Lange, 2016). This assumption plays a role in chapter 2 during my discussion of action explanation, wherein I appeal to explanatory pluralism to suggest that certain kinds of naïve causalist arguments are implausible.

3. A Final Note

There are, of course, many deep questions regarding *free* human action, rational deliberation, and moral responsibility which are inexorably linked to discussions of agency. I, for the most part, avoid these domains. This is because *before* either free action or responsibility can be philosophically considered I argue one should have a view of human action that takes agency seriously, which is to say, one that prevents the agent's disappearance. In what follows I lay the groundwork for just such an account.

CHAPTER 1: THREE NON-CAUSAL ACCOUNTS

Non-causal action theories explain actions without appealing to causation.³ My account of action in terms of essential metaphysical dependence is a kind of non-causalism. Unlike most non-causal theories, however, I explain the agent's control over her actions by appealing to an *extrinsic* metaphysical dependence relation rather than an *intrinsic* quality of the act itself. I argue that this approach avoids many of the flaws that plague other non-causal theories. In this chapter, I frame the debate by first examining three influential non-causal theories and defending them against some initial objections. The views I discuss in detail are the non-causal accounts of Ginet, McCann, and Goetz (in §1, §2, §3, respectively). I both explain and endorse their critiques of causalism while at the same time highlighting the unique difficulties each theory encounters. I do this with minimal criticism or comparison on my part so that each author receives a sufficiently charitable presentation of their own view. I conclude in §4 by discussing a crucial similarity between all of these non-causal views, which sets-up my extended critique of traditional non-causalism in chapter 2.

1. Ginet

Ginet (1990) articulates arguably the most influential non-causal account of action. He begins by noting that all acts are particular *personal* events or states, in the sense that it is *agents* who do them. Ginet formalizes such doings into a canonical statement: 'S's V-ing at t,' where *S*

³ These theories are not widely held among contemporary philosophers and, from a historical perspective, this absence is puzzling. As late as the 1950s, philosophers argued that libertarian accounts of action required "contra-causal" freedom, that is, a kind of freedom requiring the absence of any causation at all (Campbell, 1951, p. 457). Furthermore, the mid-century debates between Hempel (1942) and Dray (1957) regarding the nature of explanation also reveal the centrality of non-causal views, with Dray arguing that the "human sciences" cannot be causally explained, but only rationally explained. In like manner, action theorists like Melden (1961) argued vigorously that causal theories of action were unable to adequately provide rationalizations. After Davidson's (1963) influential arguments for causal theories of action and with the resources of indeterministic causation, brought to popular attention by Anscombe's (1971) late work, libertarians in particular no longer felt tied to contra-causal conceptions of freedom and instead embraced the promise of indeterministic causation. In the light of their long history, it's worth revisiting non-causal approaches and seeing whether a plausible version of the view can be developed. In this dissertation I focus on non-causalism regarding *action* itself, setting aside questions of free action for now.

is a person, *V-ing* is the present participle of a verb, and *t* is a specific time (p. 2). Ginet's question is thus, when is 'S's *V-ing* at *t*' an action rather than some other kind of occurrence of S's? He begins by considering and rejecting two standard causalist accounts: extrinsic causal theories and intrinsic causal theories.

The type of *extrinsic* causal theory Ginet considers is the *mental state* theory.⁴ This theory claims that S's *V-ing* at *t* is an action if and only if it (S's *V-ing* at *t*) *consists* in being caused by some combination of desire or intention and belief. This account needs clarification in at least two ways. First, the cause must happen in the right sort of way. For example, S's desire to have another drink might have *caused* S to involuntarily vomit, but it seems incorrect to say that vomiting was an action of S's. Second, the desire/belief pair must be of the appropriate sort. For example, the mental state cannot necessarily be or include a desire or intention to *V* because sometimes S's *V-ing* is unwanted or unintentional.

Even if these clarifications are met, however, Ginet argues that mental-state theories are undermined by counterexamples. For instance, if the motivational state precedes the action it causes, then there are some common actions that will not be captured by the account. As Ginet says:

Many a time, for example, I have voluntarily crossed my legs for no particular reason. No antecedent motive, no desire or purpose I expected thereby to serve, prompted me to do it . . . but it was an action if anything is." (p. 3)

Conversely, avoiding this counterexample by saying the motivational state can *accompany* an action rather than *precede* it leads to another set of counterexamples. Ginet notes that there are instances where voluntary action occurs without any accompanying intention or desire at all. For example, consider someone who thinks her right arm is paralyzed, but wants to exert herself on a

⁴ Davidson (1963) is, arguably, an early example of this type of theory.

whim. If, unbeknownst to her, her arm is not paralyzed then she has voluntarily moved her arm without intending or desiring to move her arm.

More generally, Ginet argues that there is something puzzling about this very approach to action theory. It makes something extrinsic to the personal event necessary for that event to be an action, when, all other things equal, it should be something *intrinsic* to the personal event that makes it an action.⁵ This general desideratum leads to the second kind of causalist account Ginet criticizes: intrinsic causal theories.

Intrinsic causal theories say that S's *V*-ing at *t* is an action if and only if it (i.e. S's *V*-ing at *t*) *consists* in S's causing something. To say that S's action consists in S's *causing* something is *not* the same as saying that S's *V*-ing at *t* was an action if it was *caused by* S. This account is meant to be intrinsic, thus the action is not a separate entity caused by S instead the action consists in S's causing it. This account fulfils the intrinsic desideratum, but it raises a question. What does it mean to say that S caused something? More generally, to say that X causes an effect, E, where X is an enduring thing, there must be some sort of relationship between X and E, some sort of structure that makes it proper to say that X causes E. According to Ginet, there are two ways to understand intrinsic causal theory—the *event causal* and the *agent-causal* interpretations.⁶

⁵ It is worth noting that there is no *obvious* reason why we should hold this desideratum. Plenty of things are defined extrinsically, why should actions be an exception? Perhaps Ginet is implicitly thinking that actions, because of their subjectively salient characteristics, are unique in their need of *intrinsic* characterization. I worry, however, that the more likely error theory is a simple conflation of essential and intrinsic features. Essential features are not always intrinsic (e.g. money) and for this reason our definition of action does not *have* to be intrinsic.

⁶ These event-causal and agent-causal interpretations can also be applied to *extrinsic* action theories, though Ginet does not mention this distinction. For example, *mental state* theories are a type of event-causal interpretation of an extrinsic account of action. Furthermore, one could posit that agent-causes *cause* action-events, which would be an extrinsic agent-causal interpretation. Since Ginet does not discuss these possibilities I do not examine them further.

First, on the *event-causal* interpretation, “X causes E” means X, where ‘X’ is an enduring entity, causes E if and only if there is an event C such that X is the subject of C and C causes E.

To illustrate this, Ginet gives an example of opening a door. He states:

S’s relation to that event [the door’s opening], in virtue of which S causes it, is S’s voluntarily exerting her body in connection with the door in such a way as thereby to cause the door to open. (p. 6)

This kind of relation is similar to how causation is understood more generally in nature.

Ginet argues that this analysis is incorrect. When someone pushes me into a bookshelf, knocking it over, it is true both that I am the subject of an event (hitting the bookshelf) and that this event caused an effect (the bookshelf falling). But it surely fails to count as an action of mine. An adequate account must show how the *agent* is the proper causal source rather than some deviant causal chain. A plausible way to do this is to specify a certain *kind* of neural/mental event that the agent is subject of, which underlies a volition or a “trying.”

Ginet notes, however, that this thesis has a fatal flaw. Namely, to *try* to act *is* to act. Trying is a mental action if anything is. But if this is true there is danger of a regress. For if the event of which I am the subject, which was meant to explain what makes S’s V-ing at *t* into an instance of action, is *itself* an action then there must be *another* event of which I am subject, which causes that mental event to be an action and so on. Because of this fatal flaw, Ginet turns to the other interpretation of what it means for a person to cause something.

By contrast, on the *agent-causal* interpretation, what it means for the agent to cause something is that a relation obtains *directly* between the agent as a persisting substance and the effect. There is no *event* that S is the subject of that’s therefore needed to explain S’s causing of *E*. This analysis therefore avoids the regress worry, but does so at the expense of leading to further concerns. Developing a concern first raised by C. D. Broad (1952, p. 215), Ginet argues

that an agent-causal relation does not explain all it needs to—specifically, it does not explain the time at which the action occurs. For if one side of the causal relation is an enduring agent, then what explains why the agent acts at *this* time rather than at an earlier or later time? How could the notion of something happening at a certain time (the simple mental action) be explained by something to which the notion of a certain time does not apply (the enduring agent)?

Moreover, Ginet (1990) argues that if agent-causalism were correct, then a simple mental act, like forming a volition, would have *within* it two features: the mental *event*, which is the effect of the agent-cause and the causal relation to the agent herself (pp.12-13). The *act* consists of both of these features. Ginet thinks, however, that there is no such complex structure to a simple mental act. We know the difference between the simple mental act of thinking of the number 5 and the mental event of having the number 5 come to the mind's eye unbidden. These are different events, but what makes the difference? It is not anything *external* to the mental events but rather a difference in their intrinsic quality. It is not as though both events are identical with only the agent-causal relation, conceived of as separate from its mental effect, making the difference. The agent's knowledge of the difference between these two events does not come from knowing that one is caused in a certain sort of way. Instead, it comes from something intrinsic to the events.

In the light of these criticisms, Ginet argues we should reject the claim that for S's V-ing at *t* to be an action it must consist in S *causing* something. One of the lessons of the preceding arguments, Ginet thinks, is that whatever makes an instance of S's V-ing at *t* an action (where S's V-ing at *t* is a simple mental event), it must be *intrinsic* to that event. With this in mind, consider again the simple mental act of visualizing the number 5 as compared to the simple mental event of the number 5 coming to one's mind unbidden. What *intrinsically* makes the difference

between these two mental events? Ginet claims “the mental act differs from the passive mental occurrence *intrinsically*. The mental act has what we may call (for lack of a better term) an *actish* phenomenal quality.” This *actish* phenomenal quality is the quality of it seeming to the agent that they did, or brought about, the simple mental event directly. Moreover, this phenomenal quality is all that intrinsically distinguishes such simple mental actions from other non-actional mental events. As such, Ginet concludes, it is the best analysis of action.

As he clarifies in later work, Ginet does not claim that all action is uncaused. Rather, he claims that an event is an action by virtue of non-causal factors, namely the actish phenomenal property that all simple mental actions possess. Furthermore, he claims that since all action begins with a volition, all action either *is* or *begins with* a simple mental action. Hence, he claims that:

S’s *V-ing* at *t* designates an action if and only if either (i) it designates a simple mental occurrence that had the actish phenomenal quality or (ii) it designates an event consisting in somethings’s being caused by a simple mental occurrence with the actish phenomenal quality. (p. 20)

Although Ginet does not claim that all action *is* or *must* be uncaused, he does think that in order for an action to be *free* it must be uncaused (2007; 2014). But this does not rule out unfree actions from still counting as *actions*. Ginet (2002) states, “I have *not* argued—and I see no good reason to believe—that these sufficient conditions [for an event counting as an action] *rule out* the possibility that the actions was caused either by factors that include the intention or desire cited in the reasons explanation or by something else” (p. 403). Thus, the following picture of action emerges: an event *counts* as an action because it intrinsically possesses the actish phenomenal quality—its seeming to the agent as if she directly has done it—or it is caused by an event possessing this quality. But, this is compatible with the action being causally produced. On

Ginet's account although we would still act in a completely causally determined world, our actions would not be free.

2. McCann

McCann (1998, 2012) argues that all action is grounded in the basic and simple mental activity of volition, which is itself *intrinsically* practical.⁷ Furthermore, he argues that to be intrinsically practical such simple mental acts are *essentially* intrinsically active and—on these grounds—they are experienced by us as uncaused. McCann argues for these claims by exploring the relationships between decisions, intentions, and what it means to act *for* a reason. This argument involves two stages: First, he argues that since our experience of deciding is *intrinsically* practical then, insofar as we take this experience to accurately represent its *nature*, we can justifiably believe our actions are uncaused. Second, he argues that the challenges for such non-causal theories are on par with the difficulties for causal theories, hence non-causal theories are not any worse off. McCann concludes that though these arguments are not decisive, they do show that we are justified in believing our actions are uncaused.

McCann argues for this first claim—that deliberation is intrinsically practical—by considering a typical instance of practical reasoning written as a syllogism:

(1) Would that I have an enjoyable trip to the airport.

⁷ As we saw in Ginet's analysis, a simple action is an act with no internal causal structure. In contrast, a basic action is an act that is done not by doing another action. These distinctions are different, but they coincide extensionally in McCann and Ginet's work. Though McCann mostly speaks in terms of basic actions I use Ginet's terminology of "simple mental acts" to maintain consistency, since it is clear that on McCann's view mental acts like volitions are *both* basic and simple. An interesting further question is whether *causal complexity* is the only kind of complexity that matters when discussing simple vs. complex actions. I do not here take up this question in any detail but I do think that, on McCann's account, there might be other kinds of relations that could make actions complex. For example, assume McCann's conclusion that essentially intrinsically practical actions are uncaused. Now, consider the mental act of reaching a novel conclusion through a complicated logical syllogism. To my ear, it would be strange to say that we are passive with respect to mental occurrences like "drawing an inference" or "working a proof", therefore such mental occurrences are to be counted as actions. If so, then the mental act of "reaching a novel conclusion" is a complex mental action comprised of many acts of "drawing an inference." But, if we are assuming McCann's conclusion, then all of these mental acts are essentially intrinsically practical and thus uncaused. If so, then on McCann's view it is possible to have a complex action where that complexity does not consist in *causal* complexity but instead in complex *rational* relations.

(2) I will have an enjoyable trip if I go by the rural route.
Therefore:

(3) I shall go to the airport by the rural route

According to McCann (2012) it is a mistake to understand such a syllogism as a theoretical process of reasoning *about* how to get to the airport that then *becomes* practical by the output becoming an intention (p. 247). McCann’s central suggestion is that when we engage in practical reasoning leading to an action or an intention to act, the premises are not *about* our own mental states. It is the premises’ *content* that matters. Speaking metaphorically, McCann (2012) says that our practical reasoning takes places *within* those mental states (rather than about them) and thus proceeds in terms of their content, not in terms of the events themselves (p. 248). The major premise is the *content of the agent’s desire*, its optative element, not merely a propositional description of the agent’s desiring mental state.

As such, McCann (2012) argues that “decision making is itself a matter of reasoning on my part, not just a movement of the mind that has reasoning in its background” (p. 251). Put differently, deciding is *intrinsically* reasoning and does not make reference merely to the occurrent mental states that uphold reasons, which he also calls *reason-states*. For this reason—namely, that decision-making is intrinsically a matter of reasoning—McCann argues that we are justified in believing that decision making is an uncaused act.⁸ This follows because, McCann argues, from the agent’s intrinsically practical perspective, causation is not relevant for explaining the movement from reasons to intention through deciding. The movement between the *content* of reasons to the forming of an *intention* is not, at least at first glance, a *causal*

⁸ The sense of “reasoning” McCann uses here does not require detailed deliberative processes. Even if we decide to do something without any explicitly occurrent reasoning process we still act *for* reasons and those reasons explain our decision. For more on this and how it relates to the problem of “acting on a whim” for rationalists views of action like McCann’s, see chapter 3, §5, fn 33 of this dissertation.

relation. Put roughly, McCann (2012) says to speak of deciding as caused is to suggest that it is brought about but from the agent's perspective deciding is *not* something brought about; it is something they *do* (p. 251).⁹

This sense of the deciding as a *doing* rather than something brought about, which is to say as *active* rather than *passive*, is not only intrinsic but also *essential* to the act of deciding. As McCann (1998) says, deciding is “by its own nature a purposive exercise of voluntariness—something that is essentially action in the fullest sense” (p. 173). It is conceptually impossible, McCann says, to decide involuntarily. When we decide we mean to decide and our decisions are within our control. Hence, according to McCann (1998) “there are no decisions that fail to be exercises of agency . . . none that lack the *phenomenal* character that leads us to think they are *our* doing, something we control” (p. 173).¹⁰

At first glance McCann's initial conclusion is too hasty. At most he shows that causation is irrelevant to our act of deciding, however, this does not thereby show that our deciding is uncaused. As Ginet pointed out, even if non-causal features make it the case that personal events are actions that does not preclude them from being caused. The fact that they are caused might

⁹ As I discuss below, there may be reasons to doubt this move from our experience to metaphysical claims about causation. However, I give a note of interpretive caution from the outset: McCann's fundamental perspective is one that takes our so-called practical experience “seriously”, where this means (I believe) a kind of *presumption* in favor of realism regarding the validity of our experience. This presumption is, mostly, unargued for by McCann. Regardless of what, ultimately, we think of his arguments it is important that we recognize when McCann ignores a worry not because he didn't think of it, but because he found it to be a methodological non-starter.

¹⁰ According to McCann (1998) in many of the instances where we think of our deciding as “compelled” we should understand that this compulsion is completely rational *not* nomic or causal. If I put a gun to your beloved pet bunny and then ask you to give me your jewels it is not a matter of “force” in the literal sense. Instead, I am using the persuasiveness of reasons. I am dramatically shifting your options for what counts as the most rational choice in that situation. McCann thinks that when you call such an act “involuntary” what you mean by this was that the available reasons presented to you were out of your control and had it been your choice the reasons would have been different. We do *not* mean that the decision *itself* was involuntary, after all, you could have chosen to trade the bunny's life for your money. McCann goes even further to suggest that many instances of psychological compulsion fit this model as well. On his view, psychological compulsion, like kleptomania, is a situation where in deliberating *one option* overshadows all the others as the most rational or valuable. Again, it is not that the *deciding* is itself forced, only that circumstances outside of the agents control limit the available reasons for that voluntary choice. This is very similar to Goetz's (2008) discussion of teleological determinism (p. 9).

just be irrelevant to their status as an act. McCann is aware of this concern and addresses it by emphasizing the *essentially* active nature of these intrinsically practical acts. His argument depends on two claims: (i) that our experience of simple mental acts, like deciding, is a good guide to their *essential* nature; and (ii) that there is a deep conflict between our conception of causation and the *essentially* active nature of mental acts. If both claims are true, then claiming a simple mental act could both be caused and also be an action *by virtue* of some non-causal feature, would be the same as claiming that an act could have a conflicting essence, which is absurd.

The first claim—that our experience of simple mental acts, like deciding, is a good guide to their essential nature—is not argued extensively for in McCann. He does, however, provide some reasons why we should take the practical perspective seriously. McCann (1998) argues that we only have a conception of what we are trying to explain in action by reference to our *experience* of deciding. In attempting to explain such mental acts without experiential guidance, as in a theoretical syllogism, one is unable to capture the very phenomenon of deciding.¹¹ He supports this claim by first noting that the concept of acting for a reason and acting intentionally are often interchangeable (p. 148). This is because our reasons for acting are also, in most cases, our intentions for so acting (p. 156). McCann speculates that this is why many philosophers suggest that intending simply *consists* in the mental states of desire and belief that we cite in our reasons (p. 149). McCann argues, however, that there is a significant difference between intending to do something and merely having a reason to do that thing. To highlight this, he

¹¹ This argument moves, in a compressed fashion, from the agent's perspective regarding the act of deciding to a claim about the intrinsic nature of this action. McCann, like Ginet, takes seriously the intrinsic perspective of the agent. McCann's argument, however, is also quite dissimilar from Ginet's in the strength of its claims. Ginet concludes that whatever *makes an event an action* cannot be a cause, but he does not conclude that actions must thereby be uncaused. In contrast, as we shall see below, McCann is arguing for a stronger conclusion, namely that simple mental acts, like decisions, must be uncaused if they are to *count* as action, at least in our traditional sense of action.

examines the act of *deciding* whereby, paradigmatically, reasons are made into intentions.¹²

Again, consider the practical syllogism discussed above:

- (1) Would that I have an enjoyable trip to the airport.
- (2) I will have an enjoyable trip if I go by the rural route.
- Therefore:
- (3) I shall go to the airport by the rural route.

Now, compare this to a syllogism that cites the *reason-states* rather than the content of those states:

- (1) I desire to have an enjoyable trip to the airport.
- (2) I believe that I will have an enjoyable trip if I go by the rural route.
- Therefore:
- (3) I shall go to the airport by the rural route.

According to McCann, the first *practical* syllogism more closely approximates our experience of deciding. This is because the second syllogism could, in principle, model a form of reasoning that never issues in an intention. I could theoretically *describe* my mental states or descriptively *predict* my actions using the second ‘reason-state’ syllogism but neither of those theoretical results are instances of “deciding.” As McCann (1998) says, what is reported in premise (1) of the reason-state syllogism is merely a fact about my psychology—that I have a certain desire; this is to not yet endorse anything as *worth doing*. Hence, it cannot be an accurate description of *practical* deliberation (p. 157).

In contrast, part of what makes the first syllogism *intrinsically* practical is that it does not just descriptively report the states that the agent has. Rather, it attempts to capture both the *optative* (i.e. the hopeful, wishful, and motivating *modes* of these states) and *propositional* (i.e.

¹² McCann (1998) thinks that deciding is ubiquitous, occurring whenever we form an intention prior to acting *and* even in many cases where we merely act suddenly without forming a prior intention. It does not imply, on his view, an extended instance of deliberation, though such instances do provide clear examples for us to come to understand the nature of such intention formation.

their rational structure) nature of these attitudes that, in fact, form our intentions. As McCann (1998) says:

The practical syllogism does not present my decision as inevitable, logical or otherwise. Rather it offers a purported *justification* for my decision, by citing the considerations I entertained in reaching it, and presenting the decision as *called for* in virtue of them . . . unlike causal explanations, therefore, this argument deals not in mental states but in their content. Its “reasons” are the optation and proposition in light of which I made my decision; its conclusion, rather than reporting my act of deciding, reports what I decided. (p. 153)

This, then, is partly why McCann thinks our experience of deciding is a good guide to the nature of such acts.

This, however, only shows that experiential reflection is a *necessary* guide to the essence of our acts, it does not yet show that it gives sufficient access to their nature. It is possible, McCann (1998) admits, for both the reason-state view and the intrinsically practical view to co-exist. After all, it could be that when I decide to go to the airport by the rural route my reason-states *cause* me to decide while the *content* of those states justifies the intention formed by that decision (p. 154). McCann attempts to resolve this question first by examining which model of explanation we utilize when we offer reasons for our decisions. Some of those arguments are discussed already above. In brief, McCann thinks our practice of reason explanation appeals to the intrinsically practical conception rather than the reason-state conception.¹³

¹³ I also think that, at least implicitly, McCann thinks our experience of mental acts is a good guide to their nature because there is nothing for them to essentially be *except* what we experience of them. He has no explicit argument for this, however, I think attention to the details of McCann’s account shows that we can construct an argument like Kripke’s (1973) famous pain arguments in philosophy of mind. In chapter 2, I consider this possibility and argue that, ultimately, this style of argumentation fails to establish non-causalists claims. Roughly, this is because even if non-causalists are right and our experience of simple mental acts is a good guide to what *constitutes* action, the question of whether or not they are *caused* is not a question of constitution. This claim is more important for other non-causal theories of action—particularly Goetz (2008)—so I refrain from examining it in detail until chapter 2, which concerns objections to non-causal theories.

This is still not decisive but McCann thinks that inadequacies with causal attempts at explanation make the non-causal view a more desirable alternative. This means that, ultimately, the focus of McCann's argument is the second claim I mentioned above.

The second claim—that there is a deep conflict between our conception of causation and the *essentially* active nature of mental acts—is grounded in his account of causation. According to McCann causation is an extrinsic relation that relates events, hence it is not able to adequately explain our *essentially intrinsic* voluntary control over simple mental actions. He gives several different arguments to support this claim. The first, and most direct, is that the phenomenology of deciding shows that our decisions are *essentially* intrinsically active and hence a causal explanation, which makes extrinsic features causally efficacious, is inadequate. McCann argues that if such a causal account were true then, assuming that our experience of simple mental acts is a good guide to their nature, we should expect our actions to be experienced as something that happens “from the outside.” This, however, is not the case.

This direct argument comports with the experiential evidence discussed above, however, it is unlikely that it would persuade a causalist. McCann (1998) thinks a better argument for his second claim is the problem of causal deviance. Since causal theories typically posit a *causal* relation between reasons and actions *without* reference to any intrinsic features of agency, then this causal relation cannot be sufficient for explaining why certain events count as actions (p. 115). For example, Davidson (1973) considers the case of a mountain climber who wants to get rid of the extra weight and danger of holding another climber on a rope and believes that loosening his grip will do this. The desire and belief so unnerve him that they cause him to loosen his grip (pp. 154-155). In this case, the desire and belief are reasons to perform the action he does *in fact* perform and they *in fact* cause him to perform that action, however, we would not

count this as a genuine case of action. Hence, causation, at least by the belief/desire model that McCann considers, is not *sufficient* for explaining what makes it the case that events are actions.¹⁴

The problem of causal deviance is even more apparent when considering the simple mental act of deciding. According to McCann (1998), if our reason-states *cause* our decisions then it should, in principle, be possible for instances of causal deviance *within* the process of intention formation (p. 159). So, to use the example from before, suppose besides having reasons to go to the airport by the rural route I also have reasons to visit Seattle with my brother (i.e. ‘I believe it would make a nice vacation for me and my brother’ and ‘I want to spend time with my brother on vacation’). Suppose now that by a wayward causal chain, the mental states, of which my desires and beliefs about vacationing in Seattle are the content, *cause* me to form the intention to go to the airport by the rural route. In such a case, McCann asks, did I go to the

¹⁴ Of course, McCann is narrowly targeting a simple desire/belief causal theory with these examples. Most causalists believe that deviance cases only show that *simple* causalism is inadequate for defining action. More sophisticated causalist responses to the deviance problem include: Mele’s (1992) account of “proximal intentions,” Schlosser (2007) discussion of “content causation,” as well as more recent discussions of “control conditions” such as Aguilar’s (2012) reliabilist account of control or Shepherd’s (2014) counterfactual dependence account of control. In chapter 2, I expand on McCann’s discussion and explore in more detail his version of the deviance argument and its plausibility. I do not, however, give questions of deviance the sustained attention that they deserve. In part, this is because my dissertation is, for the most part, presupposing the dialectic between causalists and non-causalists as a given and instead asking what a *more plausible* version of non-causalism would be. As such, certain traditional debates between causalists and non-causalists are only glossed. I should, however, say a few words about how a non-causalist might respond to some more sophisticated causal theories. I think that most attempts to avoid deviance by more carefully specifying the causal relation are prone to counter-examples. For example, I find the counter-examples raised by Sehon (2005; 1997) against Mele’s proximal intentions persuasive. Very roughly, Sehon argues that there is no principled way to distinguish between *deviant* proximal causes and genuine “proximal intentions” *without* appealing to notions that seem irreducibly non-causal (i.e. “teleological guidance”). In like manner, Schlosser (2007) argues that deviant causal chains happen when the *content* of a reason-state is not causally salient to the resulting bodily movement, thus the causalist can avoid the deviance objection by appealing to a content requirement. However, Davidson’s climber nervously let go *because* of the content of his reason-state (i.e. “ridding myself of the weight of another person is desirable”), yet this is paradigmatically a case of causal deviance. In contrast, causalist response which focus on specifying a *control* condition are more successful, however, there is nothing particularly *causal* about them (e.g. counter-factual control, or reliability conditions). If the solution to the *causal deviance* problem requires adverting to requirements that are, strictly speaking, non-causal or at least causally irrelevant this suggests, by my lights, that what matters for action’s definition might also be non-causal or causally irrelevant. The failure of causal specification for solving deviance counter-examples and the relative success of control accounts suggests that another lesson to be learned from causal deviance cases is the irreducibility of agential control to sub-agential conditions, hence my preference for an agency-first account of action.

airport *for* the reasons that spoke in favor of vacationing in Seattle? He thinks that such a proposal would be absurd, but there is nothing in the causal view to block such a scenario.

A causalist could attempt to avoid this objection by positing that we only act *for* the reason presented in our reason-states when the *content* of our decision matches the content of our reason-states. McCann says, however, that this response is almost *ad hoc*. If causation is what matters for explaining our actions, McCann asks, why should the *content* of those reason-states be involved? But even more telling, he points out that we never hear of such cases of causal deviance. With complex actions it is commonplace for causal failures (i.e. clumsiness, inattentiveness, muscle twitches, etc.) to make actions not match the reasons for acting. In contrast, to even imagine *deciding* to not track my reasons *for* deciding sounds absurd. To repeat an earlier point, decisions do not befall us. As McCann (1998) summarizes:

If the causal account of decision making were true, however, there ought to be such [causal-deviance] cases. For on that account the crucial relation in virtue of which a decision occurs for one set of reasons rather than another is an extrinsic one, grounded in a process that could in principle go awry. That there are no such cases indicates that reason explanations are not even undergirded by causal ones: that what makes it the case that I decide for the reasons I do is something intrinsic to the act of deciding. (pp. 159-160)

Again, McCann (1998) thinks that these considerations are not decisive in establishing his non-causal view (p. 170). Instead, he argues that non-causal theories of agency should be considered *on par* with causalist theories, insofar as non-causalism's problems are no more daunting than causalism's. He does admit that such parity arguments do not prove that *in fact* our actions are uncaused. McCann (2012) says, however, that *if* our decisions are caused then "in our decision making we labor under a systematic and thoroughgoing deception—that our practical lives are conducted within a mental apprehension that presents us to ourselves as the exact opposite of

what we are” (p. 261).¹⁵ Thus, insofar as non-causalism remains *at least* as plausible as causal theories, we should endorse a non-causal theory of action to avoid such a dramatic revision of our practical lives.

McCann considers several causalist responses to his claims. All of these counterarguments say that non-causalism fails to adequately explain actions. First, a common complaint is that *if* actions are uncaused then they are random and random events are outside of an agent’s control, hence they cannot adequately explain action.¹⁶ McCann (1998) says that these arguments usually assume that causation is necessary for making something nonaccidental (p. 179). McCann (1998) argues that this objection, in part, turns on a confusion between senses of “accidental” (p. 168). By ‘accidental’ we might mean the event is unforeseeable and unaccountable from an observer’s point of view. Insofar as this is true McCann concedes that being uncaused may be sufficient for something being accidental, a concern that he will return to later. However, there is a *practical* sense of accidental that is sometimes conflated with the unforeseeable conception.

In the practical sense of accidental an event is accidental just in case it is out of our control. A theory that made actions into such accidental events would, rightfully, be undesirable. McCann argues, however, that there is no reason to think that merely being uncaused implies

¹⁵ McCann claim about “deception” is ambiguous: Is it that we were deceived about the nature of agency (e.g. decisions are something radically different from what we thought) *or* is it that we were deceived about whether agency exists at all (e.g. decisions, if caused, are mere illusions and thus in a causally determined world we would not decide). In McCann’s earlier work (1998) he seems to endorse the stronger view, at least insofar as he argues that the *intrinsic actional* nature of simple mental acts is *essential* to the act of deciding and this is inconsistent with them being caused. In his later work (2012), however, he seems to have softened his view such that we would still decide even if there was a radical change in our concept of deciding. I believe in both cases the most salient point for understanding McCann’s view is that causalism would, by his lights, undermine any sense of agency that was ‘worthy of the name.’

¹⁶ This is similar to (perhaps even a species of) the so-called “Mind Argument” against indeterministic theories of free will. The classic source for this criticism is Hume (1772) when he claims that if we act by *chance* rather than causation we cannot be held responsible since we had no control (pp. 158-164).

such a lack of control. Again, McCann claims that our experience undermines the plausibility of this causalist assumption. It is absurd to say that we accidentally decide to do anything.

Furthermore, we do not control events by merely controlling the causes of events, such a view, McCann (1998) argues, would lead to a regress and no one would control anything (p. 168).

Instead, what allows us to exert control in the world is simply that some events, like deciding, are *intrinsically* exercises of control.

Instead of resting on the assumption that control *requires* causation the causalist could instead claim that when we decide to do something we decide *for* reasons and the only way to make sense of the “for” is for it to mean that reasons *cause* our decisions. According to this causalist line of thought interpreting the “because” relation in action explanation as a *causal* connection is the only way to make sense of how reasons explain actions. The *event* of coming to have a reason-state *causes* the action and that event is your reason for acting.

McCann (2012) responds to this more nuanced objection in two ways. First, he claims that it is a matter of controversy whether or not our decisions are caused.¹⁷ But, he says, we *know* the reasons for which we decide to do something. If this is true, then it cannot be the case that this knowledge is knowledge of a causal relation because then there would be no controversy. McCann (1998) puts it differently in his earlier work saying that if our reasons *did* cause our decisions “we could not know what my reason was [for deciding], since the question of whether decisions are nomically caused is at best moot” (p. 181). His reasoning here is compressed.

There are, I believe, two closely related arguments one from McCann’s (1998) earlier work and

¹⁷ Of course, most philosophers, in fact, think that decisions are caused. I suspect that McCann is thinking about the wider populace. Indeed, since by “causation” McCann almost always has nomic causation in mind (since, by his lights, agent-causation is a metaphysical non-starter) one might read him as claiming that it is a matter of controversy whether or not our actions are determined by laws of nature (either deterministically or indeterministically construed). If so, his claim is not as odd as it first sounds.

one from his later (2012) paper. The first argument that McCann (1998) mentions is, roughly, as follows:

- (1) If our reasons caused our decisions then we could not know what our reasons were for deciding
- (2) We (often) know our reasons for deciding.
Therefore:
- (3) Our reasons do not cause our decisions.

This argument is interesting but still unsatisfying since it does not explain what reasons justify the crucial conditional in premise (1). If, however, we flesh out justification for premise (1) we get the following argument:

- (1) Assuming our actions are caused, we do not have introspective access to the causes of our actions.
- (2) Reasons cause our actions.
- (3) So, we do not have introspective access to our reasons for acting [from 1,2]
- (4) BUT, we do (often) have introspective access to our reasons for acting.
Therefore:
- (5) Reasons do not cause our actions [from 1, 3, 4]

This is a better argument and it exemplifies the kind of reasoning McCann is attempting, however, it is too strong. McCann has not given us reason, yet, to think that moving from a lack of *introspective* access warrants a claim about whether reasons *in fact* cause or do not cause our actions. In his more recent article (2012), however, the argument is more carefully worded.

Roughly, he claims that:

- (1) When we act voluntarily something intentional, with intentional content, is done.¹⁸

¹⁸ McCann believes this because of his account of volitions. According to McCann volitions, or willings, are the basic and simple mental activity that stands at the source of all our overt actions. Volitions are, on his view, intrinsically *intentional*. This is because when we will actions it is not possible to do this “without both intending to will them, and intending thereby to produce the changes willed” (McCann 1998, p. 141). This is the best

- (2) Intentional content is always formed by deciding.
- (3) Decisions are always made for reasons.
- (4) SO, our intentional content is formed for reasons [from 2-3]
- (5) SO, we voluntarily act for reasons. [from 1, 4]
- (6) We (often) know our reasons for acting.
- (7) We do not know the causes of our actions.
Therefore:
- (8) Our knowledge of our reasons for acting is not knowledge of causes. [from 5-6]

McCann claims that introspective knowledge of our reasons for acting is the norm for all voluntary action, as indicated by our ability to report our reasons for acting. But, if our ability to correctly discern *what* our reasons were depends on which reason *caused* our actions then we should have knowledge of a causal relation. We do not, however, have such knowledge.

In this argument McCann recognizes that these considerations do not yet establish that our actions are uncaused. At most it shows that our *knowledge* of the reasons that explain our actions does not make reference to the notion of causation. However, it is consistent with this epistemic claim that, in fact, our reasons do *cause* our actions. Perhaps our reasons are only

explanations, thinks McCann, for why actions taken with very little deliberation are still considered intentional actions. This intrinsic intentionality is not just a feature of volitional activity, but also the other mental activities that guide our behavior (p. 141). For example, it is impossible to make a decision without also intending to decide or to attend to a step of deliberation without intending to deliberate (p. 141). This does not mean that intentions are *reduced* to a mere feature of volitional activity. According to McCann (1998) intentions as such are a kind of mental *state* rather than a mental *activity*. Of course, if *all* volitions are intrinsically intentional then, we can ask, what is the function of states of intending? For McCann (1998) such states allow us to execute complex plans and maintain states of intending across time for the sake of long-term goals, when we act from a prior intention one of the hallmarks is that we are aware of our prior intention and this very awareness presents us with the plan upon which we acted. In such cases the intrinsically intentional activity of volition does not replace our prior intention as presented in our state of intending, rather it “ratifies” it by transforming its content *into* the content of our willing (pp. 143-144). Thus, on McCann’s view when we say we have voluntarily acted unintentionally what we mean is that we have done something unintentionally but *with* an intention. In such cases the result or consequence of our action was *unintended* even though the actual act of willing *must* have had intentional content.

knowable *by virtue of* being the cause of our actions and we merely do not have introspective access to this further fact. McCann is aware of this limitation and leverages this first argument to ground his second, more decisive, argument.

McCann (2012) asks: What explains the fact that we *know* our reasons for acting, given that this knowledge is not knowledge of causal connections? He claims that, in light of the *intrinsically* practical nature of deliberation, we can conclude that we know our reasons for acting because the *content* of these deliberative mental states are our reasons. Recall that he argued for this claim earlier by contrasting an *intrinsically practical* syllogism with a *reason-state* syllogism. McCann claimed that focusing on the reason-states that are involved in our deliberation is inadequate because they could equally be understood as merely explaining descriptive or predictive claims. Capturing the distinctively *practical* nature of our deliberation requires referencing the *content* of those states because only they explain the optative mode of thought that underlies action. If this is so, then what explains our distinctive *knowledge* of our reasons is, in like manner, our epistemic access to the *content* of these mental states. Were this not the case and instead we only knew the reason-states involved in deliberation, then our knowledge would be insufficient to differentiate *practical* deliberation from mere *descriptive* claims about our deliberating.

For example, suppose knowledge of our reasons was knowledge of mental states it would follow that our knowledge of our own reasons for acting would be inferential. We would *know* that we had certain mental states, *observe* ourselves acting, and then *infer* that these mental states were *our reasons for acting*. This is, surely, a false picture of how our practical knowledge operates. We know our reasons for acting “from the inside” by virtue of their uptake in our intrinsically practical deliberative process. But, since knowledge of their unique practical

character is a feature of their content and the way it is held it must be that our knowledge of reasons is explained by knowing the *content* of mental states. Thus, McCann makes the following argument:

- (1) The *content* of our mental states are our reasons for acting.
- (2) The content of our mental states are abstract entities.
- (3) All abstract entities are causally inert.
- (4) Our reasons are causally inert. [from 1-3]
Therefore:
- (5) Our reasons cannot cause our actions.

Put conversely, most causalists believe events are what cause. Events, however, are by definition concrete not abstract. Insofar as McCann's earlier argument regarding the intrinsically practical nature of practical deliberation implies that the *content* of practical premises are what actually explains our decisions, it seems the *event* of coming to have a certain mental state is disqualified from explaining our actions.

Instead, McCann claims that the "for" relation is best understood as a teleological relation. The agent constructs this relation by making the content of her optative state into her reason for deciding by taking that content and making it the content of her intention. This operation is not causal but rather something intrinsic to practical deliberation. As McCann (2012) claims:

[It is] an exercise, as it were, in information processing: of investing the contents of a pair of occurrent thoughts with a new modality of thinking, the modality of intending, and of this being accomplished in and through a new occurrent thought, namely the agent's act of deciding. (p. 252).

This language of "modality" is meant to indicate that all thought has two elements: its content and the *manner* or mode in which the content is held. So, in believing that it is raining outside

my thought has both the content (it is raining) and the mode (belief) in which that content is held. What McCann (1998) means then is that in the activity of *deciding* we take the content of our reason-states and hold it in the new mode of *intending* (p. 89, pp. 134-136). McCann, admits, there is a necessary causal condition for the conscious occurrence of such mental states, but nothing about their mere existence explains the act of deciding itself. So much for the second, more nuanced, causalist objection.

A third causalist objection directly challenges this teleological view. Though reasons *can* teleologically explain why I did a certain course of action, they cannot explain why I performed the action for those reasons *rather than* others. Davidson (1963), in arguing for the claim that reasons are causes, highlights this kind of concern. As Davidson (1963) notes, we might have many reasons, $r_1 \dots r_n$, for a given action, A , and we might, in fact, perform that action. But the question remains what does it mean to act *for* reason r_3 rather than r_1 ? Davidson (1963) says that unless we can explain what it means to act *because* of r_3 we have not fully explained the action (pp. 691-692).

McCann (1998) thinks that while there is a legitimate concern behind this demand for contrastive explanations, it is not decisive (p. 182). He first makes the point that there is a contrastive problem for the causalists as well. McCann (1998) says that we can similarly ask why does *this* chain of causes exist *rather than* some other chain of causes (p. 182)? McCann (1998) argues causalists have no contrastive explanation for why *this world* with the causal chain that leads me to choose to A because of r_3 exists *rather than* the world with the causal chain that leads me to choose to A because of r_1 .

McCann (1998) does admit, however, that causal theories of action *are* better situated to explain the origination of action-events *in the world*. A causally structured world is a “seamless

fabric: every transformation of things, all that occurs, may be seen to emerge in a law-governed way from what already exists” (p. 184).¹⁹ As such, there are no explanatory gaps, at least within the world. In contrast, McCann’s non-causalism does have such explanatory gaps and thus, insofar as we want the world to be an intellectually comprehensible place, it is unsatisfactory. McCann (1998) says the non-causalist has no complete response to this objection (p. 186). Elsewhere, he (2012) argues that this objection can be understood as a conflict between our theoretical and practical understandings of the world (pp. 259-261). From the theoretical perspective there is no way to make sense of the origination of uncaused actions; such actions would be interruptions in an otherwise seamless causal story. Conversely, McCann says that from the practical perspective we *must* conceive of our simple mental acts as uncaused since only then would they be *essentially intrinsically active*.

McCann’s conclusion is that, on balance, we are justified in believing that our actions are uncaused. First, McCann’s arguments show that the non-causal view better fits with our experience of volitions as *essentially intrinsically active* and thus would not require dramatic conceptual revision. Second, that there is parity between causal and non-causal theories. Yes, non-causalist cannot adequately explain the origins of uncaused actions. But, causalists cannot adequately explain the origins of a causally structured world. Again, these arguments do not show—nor does McCann (2012) intend them to show—that our actions are, *in fact*, uncaused. Only that *if* our experiences of acts are a good guide to their nature they must be uncaused. As McCann (2012) repeatedly says, we might be deceived regarding our experience of action. But

¹⁹ It is worth noting that McCann (1998) seems most sympathetic with the process model of causation as advocated by Salmon (1984) (pp. 189-190). Furthermore, McCann (1991) has argued elsewhere for an occasionalist understanding of this causal process. These views contribute, though only implicitly, to his conception of our “causally structured world” as a “seamless fabric” rather than, for example, a collection of interrelated causal powers. This also features in his critiques of agent-causation, which I have not discussed.

insofar as the non-causal view best captures our pre-analytic notions of action and insofar as it is *no more* objectionable than causalist theories, we are both justified in believing it to be true and have more pre-analytic warrant to do so.

3. Goetz

Goetz (2008), like McCann, focuses on the mental act of deciding (or as he says, choosing) as the paradigmatic example of non-causally generated simple and basic action. Goetz (2008) explicitly says that his account is grounded in two “fundamental datum” of experience: (i) that as an agent we have the experience of making uncaused choices and (ii) that these choices are ultimately and irreducibly explained teleologically (p. 4). This experiential data, says Goetz (2008), serves to ground his belief in non-causal libertarianism in much the same way Plantinga argues that experiences can ground properly basic beliefs (p. 4).²⁰ As such, Goetz’s argumentative strategy is to first posit a certain ontology of powers and then show how this ontology is an adequate explanation for the two pieces of experiential data that he takes to be fundamental to agency.

He defends this kind of non-causal explanation in two ways. First, he claims that most causalist responses simply *beg the question* against the non-causalist. Second, he claims that many causalist objections are equally devastating for causal theories of action. He concludes, like McCann, that while there are no decisive proofs *for* non-causalism it is reasonable to hold this view since it is no worse off than causalist theories and it better comports with the fundamental data of our experience.

As noted above, Goetz frames his argument around a certain ontology of mental powers, which are meant to explain fundamental features of our experience of agency. According to

²⁰ For more on properly basic beliefs see Plantinga (1981). Roughly, a belief is properly basic if it is justified but not on the basis of any further propositional belief.

Goetz (2008), we can distinguish between those mental events we experience as active and those we experience as passive (p. 8). Things like choosing and focusing fall in the former category, while beliefs and desires fall in the latter category. This distinction, Goetz says (2008), is grounded in two kinds of mental properties: powers and capacities. He thinks that these properties are inherently different and each is an “ultimate ontological category” (p. 8).

These properties give rise to two distinct mental event types, the “exercising of a mental power” and the “actualization of a mental capacity” (p. 8). Like the properties themselves, these two types of events are intrinsically different from each other. Thus, any instance of the event-type “exercising of a mental power” is also intrinsically different from any instance of the event-type “actualization of a mental capacity.” It is this intrinsic difference that grounds Goetz’s (2008) non-causal explanation. As Goetz (2008) states, “because an agent’s exercising of a mental power is essentially intrinsically active, it is essentially uncaused or not produced” (p. 8). This then informs where he places ‘choosing’ in the ontology. Since choosing is an exercising of our mental power to choose, it is “intrinsically active and, thereby, essentially uncaused” (p. 9).²¹ This non-causal account of agency is pressed into service for the sake of libertarian arguments. But Goetz (2008) is quick to note that this is not *ad hoc* because the non-causal nature of choice follows from a more general ontology of powers (p. 9). It merely follows from his account of mental powers and mental capacities that no exercising of a mental power can be causally determined because no exercising of mental power can be causally produced at all. This conclusion should not be overstated. In Goetz’s account only *causal* determination is explicitly ruled out as a matter of conceptual analysis. As such, if one defines determinism, roughly, as the

²¹ Interestingly, Goetz does not speak much of “focusing,” one of his examples of an event-type that is brought about by our mental powers. I suspect that insofar as “focusing” is some sort of mental action it would also be uncaused according to Goetz.

claim that “there is at any instant exactly one physically possible future,” (van Inwagen, 1983, p. 3) or as “a kind of conditional necessity” (Kane, 1996, p. 8), then determinism is compatible with Goetz’s view. Neither of these understandings of determinism require that *causation* be the mechanism of determination and as such Goetz can allow for such determination.

As mentioned before, Goetz (2008) thinks that this ontology of causal powers can explain our experiences of agency. He argues for this by considering in detail how the first fundamental datum of experience—that our choices are uncaused—is explained by his ontology of powers. In particular, he considers how this ontology (i) explains the nature of choices and (ii) explains our epistemic access to our choices.

With respect to the first of these explanatory claims, Goetz (2008) contends that our experience of the nature of choice is not phenomenal. Mental acts, says Goetz, do not have a certain *feel* or quality, instead they consist *solely* in “being the exercising of a mental power” (p. 11). As such, it is “simply immediately apprehended or experienced by its agent as intrinsically active in nature and, thereby, as uncaused” (p. 11).²² This is a puzzling claim. Goetz wants to hold *both* that our mental actions like choosing have *no* feel and yet are “experienced” by us *as intrinsically active*. I struggle to make sense of what it would be to “experience” something that has no feel. In attempting to clarify what Goetz means by this claim it is helpful to first recognize his motivations for making this statement.

Goetz (2008) brings up non-qualitative experience to differentiate himself from Ginet (p. 11). Goetz thinks that merely saying mental acts have a certain “feel” that *makes them* intrinsically active and *distinguishes them* from passive mental events does not adequately

²² This distinction depends on one’s understanding of the scope of *qualia* and phenomenal experience. I, for one, think that it makes more sense as a matter of terminology to say that all conscious experiences are phenomenal though the might not all have *qualia*. It is unclear to me how Goetz is parsing up the terminology.

capture our experience of acting. Though we might have feelings that accompany mental acts, like my feeling tired after pondering a difficult question, these feelings are not *intrinsic* to the mental act itself and thus do not *make it* what it is. Mental acts have no feel. Apparently, Goetz is thinking that the phenomenal quality Ginet describes is akin to sensory qualities, like my experience of redness, or pain, or the note middle-C. If by “phenomenal quality” Ginet means to narrowly indicate only those things that have “felt qualities” where that “feeling” is meant to indicate sensory or quasi-sensory experiences then Goetz’s response is plausible. There are, however, good reasons to think this is not Ginet’s view.

Ginet explicitly says that the “actish phenomenal quality” is not qualitative *in the same way* as pain, seeing redness, or hearing sounds. The quality, according to Ginet (1990) has more to do with “the manner in which the [mental act] occurs in my mind and is not a distinct phenomenon” (p. 13). Ginet (1990) describes the quality as feeling “as if I directly produce” or “directly make it occur” or “directly determine” some mental action (p. 13). Clearly, the *feeling* of “directly determining” is not anything like a felt quality. As such, I believe Ginet’s view is actually akin to Goetz.²³ Why then did Goetz (2008) take such pains to differentiate his view from Ginet’s? The answer lies in his defensive claim that since mental acts have no felt quality “it is a mistake to argue against the kind of noncausal view that I am defending by claiming that a choice might be caused even though it phenomenologically *feels* uncaused” (p. 11). Goetz wants to link our experience of mental acts *directly* to his ontology of mental powers such that there is no room for a skeptic to suggest, as O’Connor (2000) says, that “seemings are not sufficient for realities” (p. 26). According to Goetz (2008) there are no “seemings” with respect to our experience of mental acts, we instead “immediately apprehend” them as intrinsically

²³ Goetz (2008) seems to realize this in an endnote where he reveals that Ginet has elsewhere indicated that the term “actish phenomenal quality” is metaphorical in nature. (p. 11, fn 8).

essentially active and thus essentially uncaused. This renders his central argument for how his ontology of powers explains our experience as follows:

- (1) If an event is essentially intrinsically active then it is essentially uncaused
- (2) We apprehend the event of exercising mental powers as intrinsically essentially active
- (3) If we apprehend the event as intrinsically essentially active, then that event *is* intrinsically essentially active.

Therefore:

- (4) The event of exercising our mental powers is essentially uncaused.

But both premises (1) and (2) require support. Why should we accept the conditional in premise (1) as true? Surely an intrinsically essentially active event could also be caused. In like manner, we might wonder *why*, according to premise (2), our immediate apprehension of the event of exercising mental powers avoids O'Connor's objection. This second question is particularly pressing if, as I argued above, Goetz's view is actually similar to Ginet's, for then the objection can just be rephrased to claim that "experiences" are not sufficient for realities. A further explanation for *why* such experiences can be trusted as veridical is needed. Goetz does not offer an explanation. I, however, believe some support can be given that is consistent with his overall view and implicitly figures in his explanations. I discuss each premise in turn.

Why should we accept the claim that if an event is essentially intrinsically active then it is essentially uncaused? The plausibility of premise (1) is connected to Goetz's other assertion that if an event is intrinsically essentially passive it is essentially caused or produced. In particular, it turns on his claim that such events are *essentially intrinsically* uncaused. These two concepts, intrinsic and essential, are distinct. Not all intrinsic properties are essential and, perhaps surprisingly, not all essential properties are intrinsic.

Recognizing this, we can see that Goetz (2008) *does* have an argument for his position, resting on two key claims. First, that instances or tokens of one's "exercising of a mental power" are *essentially* intrinsically active. Second, that a caused event, such as the actualizing of a mental capacity, is *essentially* intrinsically passive (p. 8). Regarding the second claim, Goetz (2008) in the cited passage *only* mentions actualizing a mental capacity as an example of an essentially intrinsically passive event. In Goetz (1997), however, he claims more generally that any entity that is the subject of an efficiently caused event is essentially passive with respect to that event (p. 197). We can now reconstruct an argument implicitly suggested by Goetz's claims. By Goetz's lights, to say that a choice could be caused but still intrinsically active to the subject as long as it did not possess its intrinsically active character *by virtue* of being caused amounts to saying the same event could be *both* essentially intrinsically active *and* essentially intrinsically passive. This is absurd. Goetz's claim about their essentiality shows that he thinks these events *are what they are* by virtue of this property, and to suggest that the same event could possess conflicting essential properties makes no sense. We can now present the following reductio style argument in support of premise (1):

(1*) Suppose for reductio that the event of exercising our mental powers is essentially causally produced.

(2*) If we experience the event of our exercising of a mental power as being a certain way, then it *is* that way.

(3*) We experience the event our exercising of a mental power as essentially intrinsically active.

(4*) If an event is essentially intrinsically passive to the subject then it is essentially causally produced.

(5*) SO, the event of our exercising of our mental powers is essentially intrinsically active and essentially intrinsically passive [1-4].

(6*) But a single event having conflicting essential intrinsic properties is absurd

Therefore:

(7*) The event of exercising our mental powers is *not* essentially causally produced.

This argument supports premise (1) in Goetz's original argument by showing that since the truth of the antecedent in premise (1) entails that a necessary condition for the falsity of the consequent does not hold, we are therefore justified in accepting the truth of the conditional. I think this is the best argument for supporting Goetz claim in premise (1). It must be noted, however, that the necessary and sufficient conditions I propose in premise (2*) are not explicitly held by Goetz. I believe this is a plausible extension of Goetz's claims regarding mental passivity and activity, however, it is also the weakest premise of my supporting argument. I consider the implications of this weakness in the next chapter. I now turn to support for premise (2) in Goetz original argument.

Premise (2) claims that we immediately apprehend or experience the event of exercising our mental powers as essentially active. Why think that our apprehension or experience of these events is trustworthy or veridical? As noted above, Goetz tries to avoid this question by claiming that our experiences are not phenomenal "seemings" that allow us to raise the objection that we might be deceived. But given the similarities between Ginet's account and Goetz it is far from obvious how *experiences*, even if they do not have a *felt* quality, avoid the criticism that they are a bad guide to the nature of the mental events. This premise is heavily criticized by causalists. Given that our *experience* of our mental actions as uncaused is one of the two fundamental datum of Goetz's argument, it is important for him to defend this claim.

Goetz (2008) considers two such objections, one from Dennett (1984) and one from Sellars (1966, 1962). Roughly, Dennett's argument turns on the claim that the microphysical processes that underlie our actions are opaque to us as agents. We are only aware of the "surface" level of our mental lives, if there are causes to our actions they would be beyond our

ability to introspectively access. As such, we cannot infer from our lack of awareness of causal factors to their non-existence. In like manner, Sellars famously distinguishes between the “manifest image” and the “scientific image.” In the manifest image we find our everyday awareness of things like choices, reasons, and desires. It may be true, Sellars says, that in the manifest image the notion of our choices being caused makes no sense. Just because uncaused actions make sense within the manifest image, however, we should not infer that our actions are categorically uncaused. For the truth of these claims are image specific, so while it is true that within the “manifest image” our choices are uncaused, the deeper “scientific image” posits a causally structured world even for our choices.

Goetz’s (2008) responds to both objections similarly. Regarding Dennett’s criticism, Goetz claims that his argument is not that we *fail* to experience the causes of our actions and then infer that they are uncaused. Goetz (2008) agrees with Dennett that such an argument would be unjustified, for the agent has no way of *knowing* that their actions are uncaused from the mere fact they are aware of no causal sources (p. 16). Instead, Goetz (2008) claims that individuals *know* they are choosing, and it follows that they are thus aware of exercising a mental power whose *nature* entails that it is an essentially uncaused event. That is, his argument is not a negative argument from our ignorance of causal sources to their non-existence but instead a positive argument from our awareness of choosing to the conceptual fact of their uncaused nature (p. 16).²⁴ Goetz (2008) primary response to Sellars is the same (p. 18).

But these responses do not adequately support premise (2). Goetz is arguing that both Sellars and Dennett misunderstand the scope of the experience to which he is appealing,

²⁴ It is worth noting that Goetz’s response to Dennett and Sellars lends credence to my modified argument in support of premise (1), for the positive awareness of our actively choosing only counts against causalist responses if one thinks there is something incompatible with a single event having both active and passive essential properties.

however, even if that is an accurate criticism it still does not answer *why* we should think that *any* experience is to be trusted as an accurate guide to the nature of the event experienced. Once again, I do not think Goetz provides further support. This is in keeping with his methodological claim that there are *no* non-question begging arguments for non-causalism. Instead, he takes the experience of our choices as intrinsically essentially active and thus uncaused to be *basic* in an analogous way to Plantinga's conception of how experiences can justify properly basic beliefs. Hence, most of his argumentation focuses on how non-causalist presuppositions are at least *as justified* as the causalists' assuming that you have the same experience of acting as Goetz.

These *essentialist* arguments support Goetz's first claim regarding the explanatory effectiveness of his ontology of powers—namely, that it explains the nature of choices. Next I consider his second claim—that it explains our epistemic access to our choices. Ultimately Goetz thinks that these *epistemic* argument function as support for his central essentialist claim.

Elements of Goetz's epistemic argument have already been raised in defense of premise (2) in the argument sketched above. More directly, Goetz (2008) argues that it is an epistemological feature of agents that when they perform mental actions they *know* they are performing this action *while* they are performing it. One natural way to explain this is to claim, as Goetz does, that we are aware of the choice we are currently making. Furthermore, this awareness must be of an intrinsic difference between a mental action like choosing and mere happenings or passive events. According to Goetz (2008) this presents a problem for causal theories of action, since in such theories there is nothing *intrinsic* to the mental event that distinguishes them from passive mental occurrences (p. 12).²⁵ On a causal theory the distinction between active event and passive event lies in their causal histories, but these causal histories are

²⁵ Goetz argument is similar to McCann's (1998) claims regarding our practical knowledge.

extrinsic to the event in question. As such, if causalism were true then we would know when we were choosing only by being aware of something external to the event itself (p. 12). But when we act we are aware of the choice *directly* and it is intrinsic features that distinguish it from mere passive events.

One objection is that the *intrinsic* features by which we know that we are making a simple mental action could be a phenomenal *sign* of the unique causal histories that produce our mental acts. There is nothing about the mere fact that we know we are choosing because of intrinsic features that precludes their also being caused. Goetz does not address this criticism directly. This may be because, as noted earlier, Goetz does not believe that our mental acts *have* any phenomenal feel and thus whatever intrinsic features differentiate them they cannot be mere phenomenal signs of an underlying cause.²⁶ Instead, these intrinsic features are their *active* nature. If so, then Goetz can deploy his earlier argument regarding how essentially intrinsically active events must thereby be essentially uncaused.²⁷

Having discussed both the *essentialist* argument and the *epistemic* argument in support of the *first* fundamental datum of our experience of acting, Goetz now turns to the *second* fundamental datum—that when we perform simple mental acts *for* reasons, they are explained teleologically. This is important for Goetz since, by his lights, causal explanations are not available for our acts he must provide a plausible alternative. For Goetz, like McCann, action explanation is teleological.

²⁶ Ginet (1990) considers a similar objection. He claims that the problem with making the intrinsic awareness of actions *only* a defeasible phenomenal sign of the *real* causal relation underlying our simple mental acts is that it does not adequately explain what makes the difference between veridical and non-veridical instances of that awareness. As he puts it, “what makes it the case that the impression is not illusory (when it is not)?” (p.13).

²⁷ I believe this relationship between his epistemic argument and his earlier essentialist argument is why he explicitly claims that the essentialist argument is “bolstered by an epistemological datum of mental action” (p. 12). The epistemic argument is only meant to be supportive.

Goetz (2008) argues for this by focusing on the criticism that uncaused actions are random occurrences. If he can show that by acting *for* a reason we act rationally, and that insofar as our actions are rational they are non-random, then he has bolstered the plausibility of teleological action explanation. He first considers a simple case of choosing. Suppose a businesswoman is rushing to an important meeting that will further her career, on the way she observes an assault happen in a nearby alley. She now has a choice. She can either continue on to the meeting or decide to stop and help as she is able. Let R1 stand for her reasons for doing what she believes is morally right and R2 stand for her reasons for furthering her career ambitions. Even if neither reason is *sufficient* for the occurrence of her choosing, either choice will be consistent with her current psychological state. Hence, her choosing will be non-random by virtue of being made *for* a particular reason, either R1 or R2.²⁸

An agent chooses *in order to* accomplish or bring about the purpose described in her reasons for choosing. As such, these rational explanations are *teleological*. Such explanations involve three parts: (1) conceiving of or representing in the content of a propositional attitude the *future* as including a state of affairs that is a purpose to be brought about or produced for the sake of its goodness; (2) conceiving of or representing in a belief the means to the realization or bring about of this end, where the means begin with the agent performing an action; and (3) making a choice to perform that action in order to bring about, or for the sake of, that purpose.²⁹

Goetz highlights, as does McCann, that conceiving of reasons as propositional states or mental events does not do justice to their explanatory direction. According to Goetz (2008), while causal explanations have a *past-to-present* direction of fit teleological explanation have

²⁸ Once again, there are similarities between this argument and McCann's distinction between two different senses of accidental events. Goetz is addressing the what McCann would call the *second* sense of accidental events.

²⁹ This is, roughly, Goetz's version of the intrinsically practical syllogism described by McCann.

future-to-present character (p. 20). This *future-to-present* type of explanation is better suited to explain choosing. Choosing, Goetz says, involves acting *in order to* achieve or bring about some purpose. This means, according to Goetz, that our reasons for choosing are “optative conceptual entities”, which are entities expressing the idea that the world *might be* a certain way that is good or in relationship to something else that is good (p. 20). If so then they are about some *future* state of the world or *future* value that the agent is working to realize, hence the *future-to-present* direction of explanation intrinsic to teleological accounts is better suited to making sense of choosing *for* such reasons.³⁰

Furthermore, according to Goetz, the causalist’s central motivations to doubt such teleological explanations are not persuasive. For example, he notes that many causalists justify their claim that a choice *must* be caused on the grounds that everyone is committed to the principle of universal causality—the principle that every event has a cause (p. 21). Goetz argues that this claim rests on a confusion. It is plausible to believe that nearly everyone is committed to something like universal explanation—the principle that every event has an explanation—but this does not entail universal *causal* explanations. There is no reason, Goetz claims, to believe that *every* event has a causal explanation.

³⁰ Goetz (2008) describes such reasons for choosing as reasons to “produce” or “achieve” some future state that is seen as valuable. Goetz argues that since reasons are optative conceptual entities they are *not* propositional attitudes *as such*. This is akin to McCann’s argument that reasons are the *content* of desire and belief states and not identical with the states themselves. Brewer (2009) is another important critic of such propositionalist understandings of action. Brewer (2009) approaches the topic as an ethicist and as such highlights the failure of propositionalist understandings to adequately make sense of certain kinds of valuable activities both because these activities have a future-to-present direction of fit *and* because engaging with their value is essentially non-productive (e.g. when we engage in friendship we do not aim to *produce* friendship, such a conception of friendship is wrongheaded) (pp. 12-67). As such, Brewer would say that Goetz, though correct in denying our reasons *must* be propositional in structure, still concedes too much to the propositionalist by claiming that we aim to “produce” or “achieve” some future state. By Brewer’s lights some values are not producible or achievable and are not even best understood as states of affairs to which these relations could apply. Though I do not take up such issues here, I do agree with Brewer on these points.

This teleological account of reason explanation faces two difficulties: first, it does not provide an adequate *contrastive* explanation of an agent's choice; second, it does not provide an adequate explanation for instances where we act for *no* reason. Goetz responds to the first criticism by presenting a more detailed account of the rational structure by which agents make decisions. When an agent chooses to perform action A for reason R1 there is also action B that the agent had reason R2 to perform. Goetz claims that *if* the agent performs A the *explanation* for this is R1. However, the contrastive question concerns *why* the agent performed A *rather than* B. One might appeal to another higher order reason, R3, that explains *why* the agent choose to perform A for R1 *rather than* B for R2.

For example, consider again the case of businesswoman S who can either do A (help the person being assaulted in an alley) or B (go to an important meeting). She has reasons R1, *that it is morally right*, to perform A and R2, *that it would further her career*, to perform B. If she performs A *for* R1, we might think there must be a further reason R3, *that she always do what is morally right*, which explains why she choose to perform A for R1 rather than B for R2. This however will not do. As Nagel (1986) points out, even if further reasons are given there is a point where this reason giving will have to stop lest a vicious regress begins. After all, if there is a second order reason R3 (*that she always do what is morally right*), which explains her actually acting for R1, then there must also be a second order reason R4 (*that she do whatever maximizes her immediate well-being*), which would explain why she choose B for R2 if she had. Thus, merely positing second order reasons only pushes the demand for contrastive explanations back a step.

Goetz (2008) responds to this critique considering a third order reason, R5, *doing what is all things considered most reasonable* (p.29). Goetz thinks that this reason answers the demand

for contrastive explanations. For, if we ask why *ultimately* S performed action A for R1 and R3 rather than B for R2 and R4, we can appeal to R5 and say that S believed by performing A she would be doing what was *all things considered most reasonable*. There is no danger of regress since there are no contrasting reasons to R5. A natural objection is that if S chooses B instead of A then it would still be inexplicable why S chose against R5 since it is *all things considered most reasonable*. Goetz, however, thinks that there *is* a kind of explanation. When S chooses B rather than A she still chose for reasons R2 and R4. This means that her action was not nonrational (since she did choose for R2 and R4) merely *irrational* or *akratic* (since she acted against R5). This can still ground a contrastive explanation since when asked why did S do B rather than A we can answer, because she acted akratically *for* R1 and R4.

Even if this account of teleological contrastive explanation succeeds, it still does not address the concern regarding agents who act *without* a reason. Sometimes we choose without a purpose, goal, or aim and we need some explanation for this occurrence. At first glance, the simplest way to do this is to say that such actions are *caused* either directly by the agent or by some other appropriate agent-involving event. Goetz (1988) disagrees, claiming that it is unclear whether positing a further cause actually does any explanatory work (p. 309).³¹ Goetz (1988) asks us to consider a case where an agent *causes* her action and there is neither a reason why she acted nor a sufficient condition or necessary causal chain that makes her act that way (p.309). We can ask, why did she cause her action? Goetz argues that by the causalists own criterion the causing of the act is random or chancy. However, if this is so then the positing an additional cause does not actually explain reasonless choices, it merely postpones the explanation one step further on the causal chain. Thus, a causal account is not explanatorily preferable to the

³¹ Goetz (1988) is primarily addressing other libertarian agent-causalists here, but I believe his point can be generalized.

teleological account, since either way such actions are random or chancy occurrences. Hence, the teleological account of action explanation can adequately address the most salient features of agency and is not any worse than the competing causalist theories.

Goetz (2008) concludes that non-causalism offers the best explanation for the two fundamental datum of experience—making uncaused choices and those choices being teleological explained by reasons—as such, it is a plausible explanation of human action. Goetz does still concede, like McCann, that these arguments are not a demonstrative proof of the conceptual truth of non-causalism (p. 35). Goetz thinks, however, that these arguments do provide *substantive* support and reveal that *insofar* as we have similar experiences of agency as Goetz we should consider non-causalism a viable explanation for those experiences.

4. Conclusion—An Important Similarity?

These three non-causalist views each articulate a picture of human agency that fits well with our pre-theoretical experience of action. This chapter avoids extensive criticism or comparison, however, in the next chapter (chapter 2) I examine in detail causalist objections as well as important contrasts between the three non-causalist accounts. I end by noting a similarity between all three views. Ginet, Goetz, and McCann all emphasize, though to varying degrees, that what makes an event count as an action must be something *intrinsic* to the event. Furthermore, in all three accounts actions originate in simple mental acts, like volitions. The conjunction of this intrinsicity requirement with the mental origination of action leads all three theories to, in subtly different ways, appeal to experiential data as our best guide for explaining action.

I suspect this similarity is because a clearly intrinsic feature of *mental* events is how those events are experienced by their subject(s). I do not want to overstate these parallels since each

non-causalist *uses* experiential claims in different ways; however, I do claim that it is this similar feature that makes all three accounts most vulnerable to causalist objections. In chapter 2, I bolster this claim by showing that while these accounts are more defensible than causalists believe they are still weak in this crucial respect. This, then, will set the stage for my own non-causal theory, presented in chapter 3, of action as *essential metaphysical dependence*.

CHAPTER 2: CHALLENGES FOR NON-CAUSALISM

In this chapter I criticize contemporary non-causal accounts for two, reinforcing, reasons. First, I argue that when responding to some causalist objections, non-causalists utilize defensive strategies that lead to dialectical stalemates, which are themselves undesirable. I call this the “dialectical stalemate objection.” Second, I argue that the reason these defensive strategies are unsuccessful is that non-causal accounts engender an explanatory gap, for various reasons, between their claims about internal experiential data and extrinsic features of the world. I call this the “incomprehensibility challenge,” since from the perspective of their causalist critics it makes the *mechanism* of action opaque and incomprehensible. I agree with non-causalists that contemporary causal theories of action are inadequate. But the standard positive arguments *for* non-causalism also fail or give rise to dialectical stalemates.

I argue for this in §1.1 by considering several standard causalist objections to non-causalism and non-causalist responses to these objections. I claim while most non-causal responses succeed there are still causalist objections that remain unanswered. In §1.2, I consider the debate between strong and weak non-causalists and argue that strong non-causalism does not provide adequate support for its position. In §2.1 and §2.2, I argue non-causalist responses to these arguments lead to dialectical stalemates by relying on internal experiential data that leaves explanatory gaps. Finally, I conclude by briefly canvassing desiderata for an adequate non-causal theory of action, thus laying the groundwork for my account of non-causalism as *essential metaphysical dependence* in chapter 3.

1. Causalist Objections and Non-causalist Challenges

I examine various objections to the three non-causal theories presented in chapter 1. Some of these criticisms already arose in chapter 1, but in this section I focus on the criticisms themselves. I first consider causalist objections and present the non-causalists’ replies. Second, I

examine challenges that arise *between* the non-causalists. In particular, I consider arguments between the “strong” non-causal accounts of McCann and Goetz, which claim actions must be *uncaused*, and the “weak” non-causal account of Ginet, which claims actions can be caused though they *count* as actions because of non-causal features.

1.1 Causalist Objections: Randomness, Rational Explanation, and Origination

Many causalist objections I presented in the previous chapter. There is wisdom, however, in returning to a few of these in detail. In what follows I discuss again three objections that causalists raise against non-causal theories. I aim to both highlight non-causalists’ responses that I find most persuasive, as well as indicate where I think causalist objections succeed or succeed once modified.

i. Randomness Objection

The randomness objection is traditional and takes several forms. Both McCann and Goetz claim this objection concerns explanatory adequacy. The critic claims that without causal explanations, our actions are inexplicable, random, or chaotic.³² For instance, Taylor (1963) argues “suppose that my right arm is free, according to this conception; that is, that its motions are uncaused . . . There will never be any point in asking why these motions occur, or in seeking any explanation of them, for under the conditions assumed there is no explanation. They just happen, from no causes at all” (p. 47). Here is one way to clarify this general line of thought:

- (1) An event is random only if it lacks a salient explanation
- (2) Events are saliently explained only if they are caused
- (3) Simple mental actions are uncaused events

³² This use of “randomness” is following terminology used by both non-causalists and their causalist objectors. It is worth noting that “randomness” can also be used in a highly technical manner that does not always track the ordinary usage within action theory.

Therefore:

(4) Simple mental actions are random

This line of reasoning, however, depends on premise (2), which is controversial.

A non-causalist should reject premise (2), while causation is one type of explanation, there are other kinds of explanation too, and types that are non-causal in nature, which suitably apply to events. The most obvious example is *rational* explanation. As Ginet, McCann, and Goetz note, agents act for reasons and these reasons explain their actions.³³ Why not only appeal to *rational* explanation to explain actions? Of course, perhaps leaving room for agents to perform actions for no reason at all is desirable; however, admitting that there might be action explanations that are non-rational does not undermine the non-causal theories described above. Admitting that *some* actions can be rationally explained, apart from any causal explanation, undermines the most direct form of the randomness objection.

To bolster the randomness objection a causalist might try to undermine the non-causalists' appeal to rational explanations of action. If the causalist makes a strong case that non-causal rational explanation is insufficient to explain action then they undermine the non-causalists approach and perhaps rehabilitate a form of the randomness objection.

ii. Rational Explanation Objection

The reasons-as-causes challenge to the non-causalist position has its roots in Davidson's (1963) paper that, arguably, revived causalist interpretations in action theory. Davidson (1963) claims that the primary reason for performing an intentional action is whatever pair of pro-attitude (desire) and belief *causes* the action that they also rationalize (pp. 685-686). He then asks, given that an agent may have multiple reasons, R₁, R₂, R₃, etc., for a single action, A, what

³³ Goetz points out that even causalists, like Taylor, recognize that merely saying the agent *causes* the action, though necessary, is not sufficient to explain the action. What then, we can ask, is the causation adding to the discussion?

explains that the agent acted *because* of R_1 rather than, say, R_2 ? According to Davidson appealing to rationalization alone does not adequately answer this question. If one builds this ‘because’ into the notion of justification itself, then the rationalization thesis becomes explanatorily weak, since there is no way to distinguish between the reasons *for which* the agent acted and the reasons which *merely* rationalize the action but are not acted upon (p. 691).

Davidson’s challenge has been interpreted two ways. First, it has been interpreted as a challenge to explain the *link*, if it is not a causal link, between reasons and actions. Second, it has been interpreted as a call for contrastive explanation. Davidson does not explicitly frame this argument as a request for the best contrastive explanation, however, many subsequent theorists have interpreted it in this way. Indeed, this kind of objection appears as an independently reoccurring causalist challenge to the explanatory adequacy of any non-causalist theory that appeals to rational justification as sufficient for action explanation. I examine both interpretations below, starting with the contrastive interpretation. I conclude that while non-causalists successfully address calls for contrastive explanations, the “linkage” objection is much more difficult.

To understand the contrastive interpretation of Davidson’s challenge, we first must clarify exactly what a request for contrastive explanation is asking for. Dickenson (2007) argues that Davidson’s challenge to the non-causalists is a challenge to provide the *best* explanation of an action (p. 2). In this way Davidson stands in line with the early 20th-century debates regarding rational explanation, best exemplified by the debate between Hempel (1942) and Dray (1957). As D’Oro and Sandis (2013) explain, these early 20th-century debates turn on methodological questions rather than ontological ones (pp. 16-17). The question of whether rational explanations could be assimilated to causal explanations mattered because it concerned the explanatory

methodology of the natural sciences as opposed to the “human” sciences. For Hempel there is explanatory unity to the sciences, hence we are justified in thinking that reason-explanation is a type of causal explanation. In contrast, Dray thought that the “human sciences” are distinguishable from the natural sciences precisely because they have different goals and different methodologies and in the human sciences these goals and methodologies *exclude* causal explanations. Davidson (1963) started the ontological turn that would dominate the mid-to-late 20th-century debates about action explanation by, more broadly, crafting a theory of action that sees actions as events. According to Dickenson (2007), however, Davidson’s explanatory challenge is still primarily a sophisticated development of Hempel’s causalist methodological arguments (p. 4).

Davidson’s argument depends on multiple-reasons cases, as noted above, but the point could be made with competing action cases as well. Either way the question is: given that the agent *has* competing reasons (for either the same action or different actions) what explains why the agent acted for *this* reason rather than another? These kinds of contrastive questions are common. It is often obscure, however, what *exactly* is at stake in contrastive explanations.

First, I note that contrastive explanations, as with all explanations, are context sensitive. Whether we should consider a contrastive explanation as, *in fact*, informatively adequate is partly a function of the explanatory interests of the inquirer. So then, clarifying what counts as the best explanation requires clarifying the relevant scope of the explanatory question. Given this, consider a schematic request for contrastive explanation: Why is A the case *rather than* B? I follow Dickenson’s (2007) terminology and call A the “fact” and B the “foil.” For this question to make sense there must be a salient connection between the fact and the foil. If I ask, for example, “why is it the case that Frodo is a literary protagonist *rather than* the Andromeda

galaxy being 1.1 billion light years closer?” it is hard to see this as a sensible question since there is no salient relationship between fact and foil.

What then, makes a salient connection between the fact and the foil? A traditional proposal (Ruben, 1987; Temple 1988) is that the fact and the foil are related by being incompatible with each other. If it cannot be the case that both A and B then there is some sense to asking, “why A *rather than* B?” This proposal, however, does not track our explanatory practice adequately as both Dickerson (2007) and Lipton (1991) argue. This is because, sometimes, when we ask for a contrastive explanation we are puzzled not because the fact and the foil are incompatible, but instead because the conditions for their existence seem symmetrical and so we expect them to *both* be the case, yet they are not. For example, “both me and my brother have been exercising and eating a healthy diet, why is he gaining muscle mass *rather than me?*” Dickerson (2007) notes that sometimes this connection is made obvious by “given” clauses. These are contrastive questions that take the following form: “Given that . . ., why A *rather than* B” (pp.7-8). According to Dickerson this highlights that explanatory salience in contrastive explanations rests on an inclusive disjunct: (1) finding a feature that A has and B lacks or (2) explaining why similar features did not entail similar outcomes for A and B.

Dickerson (2007) makes these claims more precise by formalizing it into two rules of contrastive explanation:

Basic Rule- In order to explain the contrastive phenomena the explanation must address the ‘foil’ by citing an explanatory item for the ‘fact’ and the absence or failure of the corresponding item for the foil.

Given Rule- ‘Given-contrastives’ (contrastive requests that have the form “given that . . . why A *rather than* B?”) are only satisfactorily explained when the information in the given clause is addressed. (pp.8-11)

With this we can see more clearly that what Davidson is asking for is a given-contrastive explanation. Given that both R_1 and R_2 are present, and can *both* rationalize action A, what explains the agent doing A for R_1 *rather than* A for R_2 ? Davidson's answer is simple; the only adequate explanation is that R_1 is the reason that caused A, which explains why *it* is cited (correctly) as the explanation for the action rather than R_2 , even though R_2 was present.

How can a non-causalist respond to this concern? First, a non-causalist might highlight that Davidson's challenge is *fundamentally* a challenge to forge the relationship between a reason and the action it is supposed to explain. Davidson's challenge gains traction because the rationalization relation is not strong enough nor specific enough to forge this relationship. It cannot explain the connection between a reason and the action it explains, since this same relation applies to many reasons that an agent might *have* even though they do not explain the action. Given this, it is important to note that many of the non-causalists are pointing to *more* than mere rationalization as the relationship between reasons and actions.

Consider the following examples taken from McCann, Ginet, and Goetz as paradigm cases of these explanatory relations. First from McCann (1998):

“Our real test for determining whether an agent decided for a supposed reason has to do not with nomic causation but with the replicative relationship between reasons and intentions noted earlier: that is, with whether the content of the reason-states in question is reflected in the agent's intentions.” (p. 156)

And again from Ginet (1990):

“Our anomic [non-causal] sufficient condition for such an explanation was (C3).

(C3) (a) Prior to V-ing, S had the intention to U, and (b) concurrently with V-ing, S remembered her prior intention and its content and intended that by this V-ing she would carry it out.” (p. 148.).

Finally, from Goetz (2008):

“Contrary to what Davidson would have us believe, the distinction between justifying [rationalizing] a choice and explaining a choice to which he has drawn our attention can be preserved when ‘because’ is understood teleologically . . . Teleological explanation of a choice makes reference to a future *telos*, goal, or end of an agent with respect to which she sees her chosen action as a means.” (pp.41-42).

All three of these non-causal accounts attempt, in different ways, to forge a connection between the reasons that explain action and the action itself. By their lights, Davidson is mistaken because he fails to recognize that there *is* a salient non-causal distinction between merely having a reason and acting for a reason. These examples show a difference between McCann and Ginet, on the one hand, and Goetz on the other hand. McCann and Ginet tend to focus on the relationship between our reasons (or at least the *content* of those reasons) and the intentions that underlie our actions. In contrast, Goetz focuses on our reasons as referencing future goals that teleologically explain the action. Either way, these accounts directly address Davidson’s challenge.

The non-causalist might press further here and challenge the adequacy of Davidson’s own account with respect to contrastive explanation. For, we might wonder what the relevant distinction is between being a rationalization and *rationalizing* and being a cause and *causing*? A given-contrastive explanation objection concerns the fact that some reasons are present while not having any role in *explaining* the agent’s action. As such, responding to the objection by saying that explanatory reasons *cause* the action they explain does not actually address the question. After all, the other reasons are *also* a potential cause and *also* possessed by the agent, in the same way that they are *also* a rationalization and *also* possessed by the agent.

As Dickenson (2007) points out, what is needed, even in the causalist account, is an explanation that explains how a reason can be present and yet *not* efficacious in producing the action (p.15). Dickenson (2007) argues that the most sensible move for a Davidsonian causalist is to appeal to variance between the *weights* or *strengths* of different reasons to explain the

causal efficacy of R_1 even though R_2 is present and *potentially* a causal explanation. But, Dickenson says, once the causalist has conceded this point the game has been surrendered to the non-causalist (pp. 19-20). After all, the non-causalist can *also* appeal to a non-causal conception of motivational strength—appealing to it as a purely non-causal property of reasons. This undermines the plausibility of the contrastive construal of Davidson's challenge since part of its appeal was its simple claim to explanatory superiority.³⁴ Instead, to be explanatorily superior, the Davidsonian causalist must also show how her account of *motivational strength* is more adequate than a non-causal alternative, which is a much more complex endeavor.

The causalist does have another avenue of interpretation. The linkage interpretation of Davidson's challenge provides a better ground for undermining non-causal accounts. Clarke and Mele among others have contended that the non-causalist response fails to adequately address this difficulty. To see why, consider the following case:

Paige's Pleasure Reading: Consider Paige, who starts to perform an action to get a book she had left in her office at work for pleasure reading. She wants to get an enjoyable book, but she has not decided which book, and she believes that all her enjoyable books were left in her office: call these reasons (desire/belief pair) R_1 . With these reasons she forms the intention to walk to her office to collect a book. It is also the case that a powerful neuroscientist, Sinead, has developed a 'neural-ray' (or N-ray) that allows her to hijack Paige's brain. Unbeknownst to Paige, as she opens to the door to her office and makes a choice to get book A for reason R_1 , Sinead fires up the N-ray so that it *causes* the neural (and subsequent motor) responses that lead to Paige's choosing of book A. From Paige's perspective, she forms the intention to selected book A, and acts on that intention, for reasons R_1 .

This is a slightly modified version of Mele's Martian manipulator cases.³⁵ The point of the case is to now ask, what *explains* Paige's action? According to the causalists, the criterions given by

³⁴ For another argument in favor of non-causal weighting of reasons, see Palmer (2016), (pp. 105-106).

³⁵ Clarke (2010) has an analogous objection involving auctioneering that I consider later.

the non-causalists would entail that Paige's reason, R_1 , non-causally explains the actions, but this is clearly false, it is Sinead's intentions that explain why Paige acted as she did.

To see why the non-causalist is vulnerable to this objection, consider again the accounts given by McCann, Ginet, and Goetz. By McCann's lights Paige has decided in the presence of certain reasons and *by deciding* to so act, since deciding is intrinsically intentional, Paige does so for the goals cited *in* the reasons. In like manner, Goetz is committed to saying that Paige's choice (1) represented the content of a propositional attitude about the future as including a state of affairs that is brought about for the sake of its goodness; (2) represented a belief that about the means to the realization or bringing about of this end; and (3) made a choice to perform that action in order to bring about that purpose. Finally, Ginet's account of rational explanation has two requirements: (1) that prior to opening the door to her office, Paige acquired a desire to obtain an enjoyable book, and (2) that concurrent with opening her door Paige remembered that desire and intended of her arm-raising that it contribute to satisfying that desire. On all three of these accounts Paige's action, apparently, meets their criterion for a sufficient non-causal explanation, but (so the objection goes) in fact they *do not* explain the action adequately since the *real* explanation for Paige's action is Sinead's use of the N-rays to produce her choice. The reasons and intentions cited in the non-causal accounts are explanatorily inert.

How can a non-causalist respond to this, more substantive, interpretation of Davidson's challenge as regarding the link between reasons and causes? There are three different responses by the non-causalists addressed earlier: Goetz claims that this kind of objection begs the question against the non-causalist. He thinks that deciding is *essentially* intrinsically active and thus uncaused by virtue of being an exercise of our mental power to choose. Hence, Goetz (2008) would say, it is not possible for Paige's choice to be *caused* by Sinead in the first place, since it

is not possible for a caused event to be a choice (p. 44). It might, of course, look like a choice externally and it might even *feel* like she is choosing to herself, but it would not be a choice. I find such “question begging” defenses to be rather unsatisfying; moreover, I argue below that this kind of defense gives rise problematic dialectical stalemates.

McCann (1998) does not address this modified kind of causalist response directly. He does reject causal theories of action, however, by presenting a similar objection to them. This can function as a *tu quoque* style response to the linkage objection. McCann argues that causal theories of rational explanation face an analogous series of counter-examples regarding deviant causal chains. Just like Clarke and Mele’s objections, McCann aims to undermine the sufficiency of causal explanation by positing conditions that meet causalist requirements for reason explanation, while clearly being unexplained by the reasons themselves.

To see how consider the following examples. First, consider Davidson’s (1973) classic climber case:

Davidson’s Climber: A mountaineer wants to rid himself of the weight of holding another climber on a rope, and knows he can do this by loosening his grip. The desire and belief are so unsettling to him that he is unnerved, which causes him to loosen his hold. Hence, he drops the other climber. (p. 78-79)

In this case even though the desire and belief *cause* an action it is not sufficient to *explain* intentional action. The action fails to be intentional precisely *because* it is caused. The causalist would, rightly, respond by claiming that the act was caused in the wrong way. As such, a causal theory just needs to specify what the *right* sort of causal pathways are to avoid this sort of counter-example.

McCann (1998) points out that the causalist has a clear diagnosis available to them, in cases like the mountain climber the subject’s reasons effect a bodily action *through* their autonomic nervous system, rather than via voluntary processes. Thus, a causalist could specify

that reasons causally explain actions *in the right way* if they go through voluntary neural processes. McCann raises two potential objections. First, the causalist needs to explain in more detail *why* it is the case that such neural pathways are the “right” pathways rather than the other pathways. Answering that question, McCann claims, reveals that there is no *causal* reason why one is preferred over the other, rather the causalist must appeal to intrinsic features of choosing, deciding, or intending to highlight why one of these pathways is the “right” sort of causation. If this is the case, however, why not appeal directly to the non-causal features of the reason states as most explanatorily salient? Second, if voluntary choosings were caused by reasons, even if it was (by stipulation) restricted to the voluntary nervous system, then it should be (in principle) possible for wayward causal chains to cause an inappropriate reason to cause an action.

For example, I have both a reason, R_1 , to eat lunch (“I’m hungry and desire food and I believe that eating my lunch will satisfy this desire”) and a reason, R_2 , to go to a music festival next month (“I want to enjoy good music and I believe that going to the festival will satisfy this desire”) thus it should be (in principle) possible for a wayward chain to make it the case that R_2 causes me to decide to eat my lunch. Notice that this does not appeal to a causal process *outside* of the voluntary nervous system. This raises two objections to causalist theories. First, the causalist is committed to the claim that R_2 *rationaly explains* my eating lunch. But, if anything, we might think that my eating lunch is rationally unexplainable. Second, if the causalists deny being committed to this absurdity, then they must explain why we never see such rational mismatches. It is, for the causalist, mysterious why our reasons match our intentions. McCann thinks these causal deviance arguments are decisive in showing that the reasons-as-causes accounts are insufficient for explaining actions. Thus, even if McCann’s view does not have an

adequate response to the linkage objection *neither* do the causalists to McCann's version of the linkage objection—*tu quoque* indeed.

Ginet (2008) responds by claiming that all this objection shows is that his non-causal requirement is not *necessary* for action explanation. It might be the case, Ginet argues, that some actions have either overdetermined explanations (both the intentions and the causal sources explain the action) or are explained by the causal connection. It does not, however, imply that a concurrent intention while acting is insufficient for the truth of the reasons-explanation (p. 236). A reason fails to explain action if it is *neither* presented as an appropriate concurrent intention to the acting agent *nor* the case that the reason caused the action (p. 237).

Clarke (2010) claims that this response does not appreciate the full force of the linkage objection. Clarke presents the following case:

Painting Auction: Laura wants to acquire a certain painting at an auction and believes that by raising her arm at the appropriate time she might enter a successful bid. Concurrently with raising her arm, Laura remembers that desire and intends of her arm-raising that it contribute to satisfying that desire. Unbeknownst to her, Damien, a powerful scientist who wants her to bid, has implanted a chip in Laura's brain that allows him to take control of the motions of her limbs and even produce in her the volitional act of deciding by manipulating her brain. In fact, it is Damien who causes her to will to actively raise her arm and makes sure that her so willing causes—in the way characteristic of action-production—her arm to rise. Laura bids on the painting.

As Clarke points out, this seems to be a case where Ginet's conditions are met but it is insufficient to explain the action. The intention that accompanies the action remains unimplemented. As noted above, Ginet's general strategy is to claim such cases are overdetermined. Clarke (2010) argues that this fails. Consider this parallel case where the agent does not even act:

Involuntary Painting Auction: Laura wants to acquire a certain painting at an auction and believes that by raising her arm at the appropriate time she might enter a successful bid. Concurrently with raising her arm, Laura remembers that desire and intends of her arm-raising that it contribute to satisfying that desire. Unbeknownst to her, Damien, a

powerful scientist who wants her to raise her arm, has implanted a chip in Laura's brain that allows him to take control of the motions of her limbs independently from any volitions. Damien causes her arm to rise. Laura bids on the painting.

Crucially, the arm's raising is *caused* by Damien without an accompanying willing on the part of Laura. Here it seems that both Ginet and Clarke agree that an action has not even occurred and, in fact, the event of her arm's raising is rationally explained by the intentions of Damien. Clarke then asks, what makes the difference between this case and the case where Laura's *willing* is caused?

Clarke (2010) argues there is no difference between these cases. If the intention is explanatorily disconnected in the second case (Involuntary Painting Auction) then it is also disconnected in the first case (Painting Auction). As he puts it, "In the [first case], unlike the rising one, the event to which [the concurrent intention] directly refers is an action. But that intention is no more connected to the explanandum event in this last case than it is in the rising one. The change in the character of the explanandum event cannot itself make the difference in whether the intention that refers demonstratively to that event figures in an explanation to it" (p. 30).

Ginet (2016) responds to Clarke by doubling down on his claim that the action is explanatorily over-determined. He clarifies a point obscured by Clarke, noting that the volition that the neuroscientist causes is not a further intention but rather the *most basic* part of Laura's action of raising her arm (p. 223). This does not directly address Clarke's concern, as Ginet himself admits, but it does lend intuitive support to Ginet's view that there is a close connection between the volition and the action that follows from this volition. Ginet, when considering the heart of Clarke's argument, claims that Clarke's reasoning is unclear (p. 223). According to Ginet, Clarke merely asserts that it is implausible to say that Laura raised her arm *in order to*

contribute to satisfying her desire to bid for the painting—an assertion that Ginet denies. Ginet (2016) speculates that Clarke might assert this because he cannot see how the desire can have anything to do with why Laura raises her arm “unless it figures in a causal explanation of the action” (p. 224). Ginet points out that this cannot be a premise in an argument *against* a non-causal account of action explanation without being blatantly question-begging.

Moreover, Ginet notes that when Clarke speaks of an action “implementing” or “carrying out” an intention he does not explain why that matters for action explanation, or even how “implement” might be different from “cause.” As he notes, “I would have thought it sufficient for [implementation] if the action accomplishes what is intended, whether or not the action is explained by the intention or desire the intention refers to” (p. 224). That is to say, the language of “implementation” is *too weak* to pick out the relationship that Clarke needs. The real question for Ginet is whether the agent performed the action in order to satisfy the desire to which their reason or intention refers and, thus, that reason or intention explains the action. Ginet’s conclusion is that Clarke has not given sufficient reason to think Laura’s action is not so explained.

This response, as stated, is unsatisfying. It is very difficult to explain *why* it is unsatisfying without merely reporting contrary intuitions. Some progress, however, can be made by focusing on what Ginet means by saying that the agent performed the action *in order to* satisfy their desire. This locution, “in order to,” indicates a link between the antecedent desire and the resulting action. But, what does this mean exactly? In what sense does the *agent* perform the action *in order to* fulfill *that* intention? I contend that Clarke’s argument remains plausible because this “in order to” cannot link intentions and actions without being a *thicker* concept than Ginet’s theory allows. To see why, I turn again to Clarke’s cases.

Consider again Clarke's involuntary painting auction case. Here, both Ginet and Clarke agree that the agent did not perform an action. Ginet (2016) points out that if Laura believes that she has raised her arm then she must have experienced her arm's moving as if she is raising it, rather than her just experiencing the arm's rising without it seeming to her as if she is doing the raising (p. 222). This experience, as if she is doing the raising, contains the simple mental act that Ginet calls a "volition." The belief that she forms on the basis of this experience *can* be false, but only if Damien's interference causes her arm to rise independently *while simultaneously* causing the volitional aspect of her experience. If Damien merely causes the arm's rising then it would not be experienced *as if* it was done by Laura. Conversely, Ginet (2016) argues that if Damien causes the arm's rising *via* causing the volition, which then causes the bodily exertion as it normally would, then Laura's belief is not false. She believes that *she* raised her arm and she did (p. 222).³⁶ Put another way, if Damien severs the link between Laura's volition and her arm's rising then, by Ginet's own lights, her arm's rising is no longer *her* action and is no longer explained by her desire or intention to so act.

With these features of the examples clearly in view, Clarke's objection becomes significantly more plausible. It depends on a parallelism between the involuntary and voluntary versions of the case. In the involuntary case, Ginet (2016) grants that the agent *has* a volition, that is to say, she has the simple mental event that is the most basic part of any action (p. 222). Moreover, this volition contributes to their experience *as of* raising their arm; however, the belief they form on the basis of this experience is false. The arm rises but the volition in no way *contributes* (causally or otherwise) to its rising. To appropriate Clarke's phrase, the arm's rising does not "implement" the volition, in this case because they are not causally connected.

³⁶ Of course, on Ginet's (2016) account Laura would have the further false belief that she acted freely, without her volition being subject to causal antecedents.

Remember, the debate concerns Ginet's account of rational explanation, in particular, how intentions (or desires) explain actions. Ginet claims that the following two conditions are sufficient for the truth of an action explanation, following Clarke (2010, p. 28) I apply them to Laura's case:

- (i.) Prior to raising her arm, Laura acquired a desire to obtain the painting; and
- (ii.) Concurrently with raising her arm, Laura remembered that desire and intended of her arm-raising that it contribute to satisfying that desire. (Ginet 1990, p. 143)

As Ginet clarifies, the intention mentioned here accompanies the action and refers to it directly and the prior desire explains the action by being referred to by the accompanying intention. All of these relations, according to Ginet (1990) are non-causal (pp. 142-143). But, we might wonder, what makes the difference between the possessed yet unimplemented *intention* in the voluntary case, and the possessed yet unimplemented *volition* in the involuntary case? Ginet agrees that in the involuntary case Laura's arm's rising is neither an action *nor* is it explained by Laura's prior intention. Why not *also* say in the voluntary case that though the volition causes the bodily movement the intention is still not *implemented* by that action? The intention is no more connected to the bodily movement in the first case than in the second. Put more directly, since the accompanying intention is related to the bodily movement *in the exact same way* in both cases, what we need is an explanation for why a change in the *volition's* relationship to the arm's rising changes the accompanying intention's explanatory status. Ginet does not provide such an explanation.

This objection can be framed from the other direction as well. If merely possessing the intention that her arm-raising contribute to the satisfaction of her desire is sufficient in the case of her volitionally caused arm raising to *explain* the action, then why not also in the involuntary case? In the involuntary case all three elements of action, according to Ginet, are present. The

intention, the volition, and the bodily movement are all there, however, the volition and the bodily movement are *causally* disconnected. The force of Clarke's objection comes from the fact that changing the link between the volition and the bodily action makes no clear difference with respect to the *intention*. By Ginet's own lights, there is no change in the relationship between the intention and the volition/bodily action in either case. Given this, an explanation is needed for *why* the mere fact that in *one* case there is a causal link between the volition and the bodily movement should *matter* with respect to the intention's contribution to *explaining* the action. How does this fact constitute the agent's performing a bodily movement *in order to* fulfill *that* intention?

These considerations show that the *linkage* objection is devastating for *weak* non-causal view, like Ginet's. Additionally, part of the strength of Clarke's position is that it is the more explanatorily parsimonious theory. Before moving on to consider the final causalist objection I offer a potential counter-example to Clarke's claim of explanatory parsimony. My aim here is to highlight that, at the end of the day, we are in a dialectical stalemate with no clear path forward. At first glance, this reduces all non-causalist strategies to those pursued by the strong non-causalist—namely, *tu quoque* style responses and accusing the causalist of question begging. Here is a final case that is meant to run parallel to Clarke's case. The force of Clarke's case in part comes from the fact that it shows that the *causal* story is the more explanatorily salient one. That is, it doesn't just show that the alternative view is insufficient it also shows that his view has explanatory teeth. For that reason, it would be useful to craft a parallel case for the non-causalist. This example turns on a distinction, already raised by McCann, between reason-states (i.e. either the mental *types* that are reasons or their physical *realizers*) and the content of reason-states (i.e. the abstract entity that specifies what the reason is *about*).

Opera Outing: Suppose reason-states can cause actions. Gillian is aware that a local community opera company is putting on an overly ambitious production of *Percival* and she is about to decide if she wants to attend. If confronted with a reason that involves pleasing her friends, Gillian is disposed to attend *Percival* despite its almost guaranteed aesthetic flaws. The reason-state she currently has regarding the opera, however, makes no mention of her friends in its content, instead it is *about* the aesthetic qualities of the opera. Her neuroscientist friend, Shannon, has developed a special device that projects *mental-rays*—beams that change the *content* of mental states without changing the states themselves. Right as Gillian is about to decide Shannon uses the *m-rays* to change the content of her reason-state to involve pleasing her friends. Gillian decides to attend the opera despite her aesthetic misgivings and her reason-state causes her to do so.

We can ask, what explains Gillian's choosing to attend the opera? By the causalists lights it is the reason-state which *caused* her to perform that action. But, this is surely false. The only reason-state on offer was causally inert *until* the content of the reason changed by virtue of Shannon's intervening. Furthermore, this is not an issue of deviant causal chains, for nothing was *causally* amiss about the processes by which the reason-state was formed and in turn caused Gillian's action. This is a case where what explained her going, rather than not, was the non-causal change in content. The causal structure remains fixed.

A causalist might protest that this case begs the question against a causalist by assuming that the content can change (non-causally) and then be causally efficacious because of that change. I believe there are two plausible responses to this concern. First, if this style of case is question begging to the causalist then it seems open for the non-causalist to claim the Clarke and Mele's analogous cases are question begging to non-causalist positions. The cases stand or fall together. Second, it is unclear to me *why* such a view would be question-begging for a causalist. I take it to be the case that most action theorists, who are not epiphenomenalists, are committed to the straightforward claim that reasons are reasons *for* or reasons *against* an action of mine precisely because of their content. As such, to deny that shifting the content of a reason-state can be efficacious *or* to deny that one can even metaphysically do this without changing the identity

of the reason state would be to smuggle in controversial thesis in philosophy of mind regarding either reduction or epiphenomenalism. Insofar as we resist making these further concessions, the causalist has an analogous problem regarding the *sufficiency* of causal action explanation.

Where do the above considerations leave non-causalism overall? Weak non-causalism is subject to counter-examples that highlight its inability to explain the *link* between reasons and actions. Strong non-causalism avoids this, but only by acquiescing to argumentative tactics that undermine the productivity of the dialectic. Strong non-causalists tend to utilize *tu quoque* style responses or claim that their causalist opponents beg the question. These are, of course, legitimate argumentative strategies, however, they lack persuasive force. They are purely defensive and mostly amount to burden shifting.³⁷ This leads to the final causalist objection I will consider, for it is an objection that cuts to the heart of even the strong non-causalist response.

iii. Origination Objection

I believe, along with McCann, that the randomness objection consists of two different concerns: (1) The agent's actions are accidental or do not have the required explanation to count as an action of hers. (2) There is no explanation for the *coming into being* of the agent's actions. They do not have a sufficient ontological explanation, thus they appear out of nowhere. I discussed (1) while discussing the randomness objection above. I left concern (2) undiscussed in part because I think it is best framed as a worry not about randomness but about *origination*. Where do actions *come from*, metaphysically speaking, if they are not part of the causal nexus? I take this language of "origination" from McCann (1998) who clearly recognizes this as a difficulty for strong non-causalists. McCann claims that his view only "address(es) the rational grounds for decisions, not their ontological underpinnings" (p. 186). Because of this, according

³⁷ Again, these are legitimate strategies. Indeed, I suspect that most of philosophy involves determining where the burden of proof lies. They are, however, also easy ways of obscuring more productive avenues in a dialectic.

to McCann, strong non-causal accounts of the will “represent discontinuities in the world, and their provenance is as hidden from us as that of the world itself” (p. 187). As noted in the previous chapter, McCann thinks that this is a reasonable critique of his style of non-causalism. He offers two responses. First, causalists are subject to a similar objection (*tu quoque*), insofar as they have no explanation for why the causal nexus exists in the first place (pp.183-190). Second, this discontinuous world *just is* the world of our experience. Thus even if we do not have a full explanation, we are justified in positing its existence as necessary for our conception of agency. That just is what it means to take the “practical perspective” seriously.

McCann himself recognizes that the first response is weaker than he would like, for even if the causalist lacks an overarching explanation it seems that the strong non-causalist has left many more entities ontologically unexplained. As such, the core of his response is to suggest that the balance of reasons is on the side of the strong non-causalist. Even though his non-causal account leaves somethings unexplained, the practical perspective is central enough to our lives to warrant holding this view despite its inelegancies.

I am unsure this argumentative gambit works. It requires sacrificing our theoretical reason for the sake of the practical. Perhaps our pre-theoretic concept of practical agency is incompatible with our theoretical attempts to wholly explain events in the world. If so, then it is open for a causalist to respond by simply saying, “so much the worse for our conception of practical agency.” In a manner similar to other strong non-causalist responses, McCann’s response though useful as a mere defensive measure is unlikely to push the dialectic forward.

Leaving the particularities of McCann’s account aside, the question of origination is troubling for other non-causal theories. In the context of *free* action, even Ginet’s weak non-causalism falls victim to a similar concern. Ginet claims that free actions are uncaused. As such,

it is puzzling how such actions come into being. Ginet (2007) contends that this supposed puzzle is itself ungrounded. He states, “it seems evident to me that, given that an action was uncaused, all its agent had to do to make it the case that she performed that action was perform it” (p. 247). Ginet (2014) has maintained this position elsewhere, further claiming that such a view is, in principle, compatible with the natural order by virtue of not causally producing but merely “limiting” the choices available to the agent (p. 25). While Ginet (2007) recognizes the origination objection, at least in the guise of a randomness objection, he contends that causalists have yet to explain adequately why such ontological “randomness” entails that actions cannot be “up to the agent” (p. 248).

At least with respect to free action, this response places Ginet’s weak non-causalism in the same camp as McCann. It is primarily a burden-shifting response that serves to, obliquely, accuse his causalist opponents of begging the question. Again, as I noted with McCann, while such a response is a useful defense I believe that it is inadequate for moving the dialectic forward. Goetz’s (2008) response to this origination concern is similar. Goetz (2008) attempts to ground his reply to the origination objection in the mental ontology that he posits, which entails that our exercising of the power to choose is *essentially intrinsically* active and therefore uncaused. He then claims that, since we know that we choose via our experience of choosing, we are justified in believing that this is an expression of our active power to choose that must therefore be uncaused. Thus, by Goetz’s lights, to answer the question *how* do uncaused choices originate we appeal to the mental power of choosing, which is in turn supported by our experience of choosing as the best explanation for the active nature of choice. Though this response does have more going for it than Ginet’s or McCann’s I think, at its heart, it is still a

“begging the question” style response. Goetz does not give us much reason to accept his mental ontology other than appealing to the *very experiences* the ontology is meant to explain.

Of course, while these responses are *thin* that does not mean that they are wholly inadequate. In section 1.3 and section 2, I offer a more complete account of why this style of response is inadequate. Furthermore, it is worth noting that while Ginet’s account gives a *thin* response to the origination objection it is still open to him as a weak non-causalist to maintain that an event is an action *by* its phenomenological non-causal features but that the existence of such events is ultimately causally explained. That is, Ginet could give up on the nomological possibility of free action if the dialectical cost of maintaining a non-causalist response to the origination objection becomes too high.

At this juncture we see a split among non-causalists. While non-causalists have many adequate responses to causalist objections, there are two criticisms that deserve more consideration: First, Ginet’s weak non-causalism is subject to the *linkage* version of the *rational explanation* objection. That is, it cannot adequately respond to counter-examples, such as those given by Mele and Clarke, which show that there is no explanatorily salient link between intentions and the actions they non-causally explain. Second, Goetz’s and McCann’s strong non-causalism is fragile with respect to the *origination* objection. Though both authors have responses, these responses are decidedly *thin* and this raises the larger concern that strong non-causalism is, ironically, dialectically weak. By relying too often on *tu quoque* and “begging the question” counter-responses McCann and Goetz hamper their ability to participate in a philosophically productive dialectic with their causalist opponents. This naturally leads us to ask, between the two non-causalist strategies—strong and weak—which is to be preferred? These debates among non-causal views are useful both for clarifying the views themselves and more

clearly understanding their plausibility in the overall debate. In the next section I turn to these *non-causalist* challenges with the aim of adjudicating between weak and strong non-causalism. Afterward, in my final section, I offer my own criticism, which synthesizes the best elements of causalist objections—the *linkage* and *origination* objections—with a larger concern about dialectical stalemates and incomprehensibility. This, then, clears the way for my own non-causal account in chapter 3.

1.2 Non-causalist Challenges: Strong vs. Weak Non-causalism

McCann and Goetz are “strong non-causalists,” they both argue that simple mental actions *must* be uncaused. This contrasts with Ginet’s “weak non-causalism”, which claims that such simple mental acts may be caused though they are not actions *by* being caused. Ginet believes that only *free* actions must be uncaused. McCann, for one, does not seem to entertain the free action/unfree action distinction. He insists that insofar as we choose at all we are free, at least from causal determination. Goetz entertains the possibility that our choosing could be teleologically determined, however, direct causal determination of our actions is not possible on his account either. Thus, for McCann and Goetz, a causally determined world would be a world without action.

I agree with Ginet that actions can exist in a world wholly causally structured they just would not be actions *by virtue of* anything causal.³⁸ I argue for this by first, taking Goetz as a paradigm example, reconstructing the strongest argument for the claim that *all* simple mental acts must be uncaused. I then defend this strong non-causalism from objections raised by Palmer (2016), which I believe are inadequate. I then raise my own objections to the view, again

³⁸ Moreover, though it is not important for the main point of this dissertation, I agree that actions must be completely uncaused if they are to be *free*.

paradigmatically held by Goetz, ultimately siding with Ginet that actions could exist in a causally determined world though they would not be free.

Goetz (2008) argues that all simple mental acts are uncaused because they are each instances of the kind “being the exercising of a mental power” and “because an agent’s exercising of a mental power is essentially intrinsically active, it is essentially uncaused or not produced” (p. 8). This, according to Goetz, follows because to be caused is to be intrinsically passive with respect to the cause. McCann (2012) echoes similar considerations, saying that “what befall me are events in which I am passive rather than active, things that happen to me instead of being done by me. None of this pertains to deciding” (p. 254). What are we to make of these arguments? At first blush they seem to be versions of the argument I constructed in the previous chapter on behalf of Goetz:

- (1) Intrinsic features of a mental event type (or kind) ontologically define it, and differentiate it from other mental event types (or kinds)
- (2) An entity cannot possess ‘contradictory’ intrinsic properties at the same time and in the same manner
- (3) If simple mental events are caused then they are intrinsically *passive* with respect to their cause.
- (4) Simple mental acts are experienced as intrinsically *active*
- (5) Simple mental acts cannot be caused [from 2-4]
- Therefore:
- (6) Simple mental actions must be uncaused [from 1 and 5]

This argument captures, roughly, the reasoning deployed by Goetz and, to a lesser extent, McCann; however, it fails to prove its conclusion. Premise 3 and premise 1 are the two weakest premises in the argument and ultimately, I argue, that premise 3 is false. But it is tempting to dismiss this strong non-causalist argument too quickly. To understand, fully, the reasoning behind this position I turn first to an inadequate critique and defend Goetz’s version of strong non-causalism against this critique.

Palmer (2016) argues that Goetz non-causal view is inadequate, and instead a revised non-causal theory should leave space for choices to be caused even if *free* choices must ultimately be uncaused. Palmer's critique rests on two considerations: First, Palmer challenges Goetz assumption that choices must be uncaused. By contrast, Palmer thinks that while it might be necessary for *free* choices to be uncaused it is not clear that all choices have this constraint. Second, Goetz's view, according to Palmer, leads to three implausible consequences, (i) it follows that all choices are *ipso facto* free; (ii.) it follows that causal determinism rules out any kind of choice whatsoever, and (iii.) it follows if causal determinism is true then deliberation can never be successfully concluded by way of choices.

While, as I noted above, I ultimately agree with Palmer's conclusion I contend that his argument against Goetz has errors that obscure the important underlying issues at stake in the non-causalist debate. I focus primarily on Palmer's first concern and the argument he uses to support it. This is because I believe it is the most interesting and substantive argument and whatever plausibility Palmer's second set of concerns has comes from the success of his first argument.³⁹

Palmer's first claim, that actions do not *have* to be uncaused, rests on the following case. Consider Jane who is deliberating about whether to stay at home or go to the movies. She has

³⁹ If one assumes, as I eventually show, that Palmer's first challenge is ineffective then the second set of concerns that Palmer raises can be briefly dismissed: (i) it *does not* follow from Goetz's account that choices are *ipso facto* free, Goetz (2008) only argues that they cannot be caused but as he carefully points out there are, for example, other kinds of determination. In fact, though he does not discuss this possibility directly (and I'm sure he would think it false) it is at least consistent with his view that our choices could be teleologically determined *by virtue of* the physical universe being causally determined, where this "by virtue of" indicates some sort of non-causal grounding relation. (ii) that causal determinism rules out any kind of choice and (iii) that if causal determinism is true then deliberation can never be successfully concluded by way of choice are both accurate descriptions of what is entailed by Goetz's position, but it is unclear to me how these claims are supposed to undermine his argument. It strikes me that Palmer's concerns in (ii) and (iii) are both examples of burden shifting. Of course, burden shifting can be justified, but I take it that what justifies it in this case is the success of Palmer's *first* challenge, which I think is, in fact, inadequate as argued above.

both the desire to stay at home and a desire to go to the movies. After some deliberation she chooses to go to the movies. By stipulation her desire to go to the movies *causes* her to make that choice (p. 100). Palmer claims that in this case it is almost undeniable that Jane has made a choice. Perhaps it could be argued that her choice, because it was caused, was unfree and thus she should not be held responsible for it. But, says Palmer, this surely does not undermine the simple claim that she made a choice.

This is not yet an argument. It is a judgement in a case in which Palmer stipulates that event of her choosing can be caused. Though, I take it, Palmer hopes that this case primes our intuitions to question the strong non-causal view. To support these nascent intuitions Palmer offers further support for his considered judgement. He notes that from Jane's perspective she certainly believes she is choosing and one could plausibly imagine that if we brought to her attention that her choices was caused she would still think she had chosen. Though this reaction is natural Palmer notes that Goetz might respond that Jane is just mistaken. Instead, Goetz would highlight the conceptual foundations of his argument noting that according to him it is a *conceptual* truth that choosing is an essentially intrinsically active event. Further, Goetz states that events which are efficiently caused are produced by that cause and as such they are occurrences that a subject is passive towards. To this Palmer responds:

This claim is not self-evident and Goetz provides no support for it whatsoever. Goetz does argue that choices are *intrinsically* active, which follows from his claim that choices are the exercise of a of a person's mental power to choose, and a person's exercising a mental power is an intrinsically active event . . . perhaps it follows that if something is intrinsically active, then it cannot derive its active character from its relationship to any antecedent event . . . [but] it does not follow that this thing cannot be caused. (p. 101).

Though the active nature of choosing must be intrinsic to the action it is, by Palmer's lights, still completely open that there are other extrinsic relations to which the event may be subject. Thus,

Goetz has given us no reason to accept his view, especially considering the intuitively plausible understanding of Jane's case wherein she chooses though perhaps not freely.

I think the best argument we can attribute to Goetz is more nuanced than Palmer gives him credit for, though I admit that Goetz's reasoning is quite compressed. The main problem with Palmer's critique is that he glosses over a key concept that Goetz deploys in his description of our experience of choosing. Palmer frequently cites Goetz (1998) as saying that our choosing is "intrinsically active." This is true but incomplete. Throughout, Goetz (2008; 1997) states that our choosing, and more broadly *any* mental event that counts as an exercising of a mental power, is an *essentially* intrinsically active event. This difference is subtle but important for the plausibility of Goetz's overall argument. To make sense of Goetz's argument, we must first understand how these terms are being used and their possible combinations.

Intrinsic properties are generally understood to be those properties that are "proper" to the entity. As Lewis (1983) states, "a thing has its intrinsic properties in virtue of the way that thing itself, and nothing else, is. Not so for extrinsic properties, though a thing may well have these in virtue of the way some larger whole is" (p. 111). This is sometimes mistakenly conflated with essential properties. Essential properties, as opposed to accidental properties, are generally understood as those properties that an entity *must* have, where "must" indicates a kind of modal necessity. In contrast, accidental properties may or may not be predicated of an entity, they are merely modally possible. Though this view is the dominate one, it is worth noting that there are compelling arguments from Fine (1995a, 1995b, 1994) among others that suggest that a modal account of essential properties is too weak to sufficiently characterize the phenomena.⁴⁰

⁴⁰ For example, it entails that my property of "being such that if I exist there are infinitely many primes" is an *essential* property of myself. This seems strange for that property has nothing to do with *me* at all. Fine (1995a) advocates supplementing the modal definition with something more directly related to the entity under consideration. In particular, Fine suggests that it would be better for essential properties define the entity. This is

Regardless of the proper characterization of essential properties the most important features for my purposes is (i) that they can be distinct from intrinsic properties and (ii) they characterize the entity that they predicate.

As I mentioned in chapter 1, once we recognize this we can see that Goetz (2008) *does* have a kind of argument for his position. For he makes two claims. First, that instances or tokens of one's "exercising of a mental power" are *essentially* intrinsically active. Second, that an event that is caused, such as the actualizing of a mental capacity, is *essentially* intrinsically passive (p. 8).⁴¹ We can now see how Palmer's attempted counter-argument falls short. Again, as I noted in Chapter 1, Palmer's claim that a choice could be caused but still intrinsically active as long as it did not possess its intrinsically active character *by virtue* of being caused, amounts to saying the same event could be *both* essentially intrinsically active *and* essentially intrinsically passive. According to Goetz, this is absurd. Goetz thinks these events *are what they are* by virtue of this property, and to suggest that the same event could possess contradictory essential properties makes no sense.

Palmer might respond that Goetz has given us no reason to agree with his claim that caused events are *essentially* intrinsically passive and thus he has given no reason for think that an event cannot be both caused and intrinsically active in character. This response is effective because Goetz does fail to provide a clear argument for this claim. I do think, however, that there are enough indications in the text for us to recognize and reconstruct a kind of argument.

puzzling for those who think of definition as primarily about words or concepts not objects or other natural entities. It is likely that Fine is appealing to an older conception of definition such as that, plausibly, advocated by Aristotle in *Metaphysics*. I do not, here, take up the task of explicating this alternative conception of essence, though I note that I find it plausible.

⁴¹ In the cited passage from Goetz (2008) he is taking explicitly about actualizing a mental capacity as an example of an essentially intrinsically passive event. In Goetz (1997), however, he also mentions more generally that any entity that is the subject of an efficiently caused event is *essentially* passive with respect to that event (p. 197).

First, it is noteworthy that Goetz allows that an event might be *dependent* on antecedent factors and that we can distinguish between this dependence relation and the active character of the event. As Goetz (1997) states, “while it is true that a choice occurs in the context of and is dependent upon the occurrence of antecedent events (e.g., the coming to have reasons to act and thinking about means to an end), it is plausible to think that a choice does not derive its active character from these events” (p. 198). This is notable because, as Palmer (2016) argues, such a position is compatible with the claim that active events are caused since an active event could be caused without deriving its active character from that cause (p. 101). We should charitably speculate that since Goetz recognizes this distinction he also recognizes its compatibility with causal production, as such it must be something about the *nature* of the event’s active character that led him to the conclusion that a mental event like choosing *cannot* be caused.

Goetz is unpersuasive here. He claims it is a conceptual truth that the exercising of a mental power is essentially uncaused and that this is a something we *know* because of our experience of the essentially intrinsically active nature of choosing. I do not see how these considerations support Goetz’s essentialist claims. I think, however, that there is a nearby argument that Goetz could deploy that would better support his contention. Furthermore, as I alluded in the previous chapter, this nearby argument bolsters McCann’s view as well. As such, I present it in some detail though it is more of a friendly addition to Goetz’s (and McCann’s) arguments, rather than a strict explication of his ideas.

Why might our experience of choosing as *intrinsically* active gives us reason to make the further claim that they are *essentially* intrinsically active? The answer lies in *mental* events. The argument is analogous to Kripke’s argument regarding identity theories in philosophy of mind. According to Kripke (1972), there is an important disanalogy between the following two types of

identity claims: (i.) heat is molecular motion and (ii.) pain is c-fibers firing. Kripke argues these kind of identity claims are necessary. Once we name a natural kind we rigidly designate the entity under discussion. Certain elements of our sense experience might serve to *fix* the entity we are rigidly designating, but these descriptive properties are not to be understood as the entity itself or as an essential property of that entity. We can, therefore, give an error theory for why someone might mistakenly think the heat might *not* have been molecular motion. As Kripke (1972) says, “When someone says that heat might have turned out not to be molecular motion, what is true in what he says is that someone could have sensed a phenomenon in the same way we sense heat, that is, feels it by means of its production of the sensation we call ‘the sensation of heat’ . . . even though that phenomenon was not molecular motion” (p. 150). The converse is also true. Given the necessity of identity, what someone means by saying “heat might not have been molecular motion” is something more like “creatures might have been such that they did not *feel* the phenomenon of molecular motion (that is, heat) by experiencing the sensation of heat.” Put briefly, the sensation and the *entity* that the sensation indicates are two different things, heat is the name we just so happen to give to the entity that it was fixed by the contingent *sensation* of heat.

In contrast, the identity relation of pain does not seem to admit to this kind of error theory. If pain is c-fibers firing, and we wanted to explain the possibility of c-fibers firing without us *feeling* the sensation of pain we would be at a loss. Pain *just is* the phenomenal experience in a way in which “heat” is not. The sensation does not merely fix some other entity, the sensation *is* the entity. Again, as Kripke states, “in the case of molecular motion and heat there is something, namely, the sensation of heat, which is an intermediary between the external phenomenon and the observer. In the mental-physical case no such intermediary is possible,

since here the physical phenomenon is supposed to be identical with the internal phenomenon itself.” Pain is picked out by its own *essential* property, the property of being pain itself. As such, if it is even possible for c-fibers to fire without someone having the sensation of pain then pain must not be identical with c-fiber firing.

I believe that a similar argument supports Goetz’s (2008) claim, implicit in premise (1.), that our experience of mental activity is a good guide to their nature. Though he does not think that our choosing has a *felt quality* he does think we *experience* it as active. Consider again his statement that:

Though a choice, unlike an experience of pain, lacks any qualitative feel, I think the view that it too has a nonrelational, intrinsically active nature that is introspectively accessible to its agent is bolstered by an epistemological datum of mental action . . . If [the agent] knows in this way that he is choosing, then it seems to follow that choosing is intrinsically different from a mere happening or passive event *and that he is aware of this difference*. (p. 12, my emphasis).

Any other attempt to explain such simple mental actions other than the agent’s direct experience of choosing directs attention away from the very event being explained. There is nothing for the simple mental event of choosing to be *other than* our experience of it as an intrinsically active event, hence it is also essential for it to be an intrinsically active event. Just like pain, the simple mental act of choosing is epistemically picked out by its own essential property of “being intrinsically active.” The same reasoning would apply, *mutatis mutandis*, to the intrinsically passive simple mental acts, which are caused.

If my generous reconstruction of Goetz argument is plausible, then it seems he *does* have an argument for why we should think that our experience of choosing is *essentially* intrinsically active and thus *must be* uncaused, because caused mental states are *essentially* intrinsically passive, and the same mental even cannot possess conflicting essential intrinsic properties. Hence, Palmer’s criticism is insufficient to address Goetz argument.

How then do we adjudicate between the strong non-causalism of Goetz or McCann and the weak non-causalism of Ginet? Though Palmer's critique is inadequate I think there is another path to rejecting strong non-causalist arguments. Though my reconstruction of Goetz's argument does more than merely assert that our actions are intrinsically active, it still has several flaws.

Recall the modified argument given in chapter 1:

1. Intrinsic features of a mental event type (or kind) ontologically define it, and differentiate it from other mental event types (or kinds)
2. An entity cannot possess 'contradictory' intrinsic properties at the same time and in the same manner
3. If simple mental events are caused then they are intrinsically *passive* with respect to their cause.
4. Simple mental acts are experienced as intrinsically *active*
5. So, simple mental acts are uncaused [from 2-4].
Therefore:
6. Simple mental actions are uncaused [from 1 and 5].

As we have seen, Goetz has given us some arguments for why we should believe that mental events, like choosing, possess their essential properties intrinsically and that these essentially characterize them. Further, once this is accepted, it is clear how Goetz moves from the intrinsic *passivity* of caused mental events, through the claim that a single entity cannot possess 'contradictory' intrinsic properties, to the conclusion that choosing must be uncaused. But though we now understand the plausibility of Goetz's argument—in particular premise (1)—we have not yet considered the plausibility of his crucial premise (3), namely, "if simple mental events are caused then they are intrinsically passive with respect to their cause." What recommends the conclusion that caused mental events are *essentially* intrinsically passive? I suggested above that one possibility for Goetz is to appeal to parallel reasoning as that for active mental events using Kripkian style argument. This strategy, however, fails to provide adequate support for premise (3).

To see why, consider again how Kripke argues for the claim that the property of pain which is phenomenally accessible to us is *also* the essential property of pain. It is because he is arguing with a type-type theorist that this kind of move is available to him. Because *identity* relations are necessary, conjoined with the fact that pain could be nothing other than its phenomenal features, we can conclude that the phenomenal features we experience are *also* its essential features. A crucial element of the argument is that it is a claim about *identity*. The type-type theorist is saying that pain *is* c-fibers firing, however, if that is true then that theorist needs to explain the seemingly contingent relationship between the phenomena and the material substrate of the phenomenon. This makes the question of pain very different from the question of simple mental actions.

The causalist is *not* asking constitutive questions or a question of identity. To ask, “are actions caused” is *not* the same as asking, “are actions identical with effects.” To see why consider pain again from a Kripkian perspective. Pain is, as far as we know, caused. When I accidentally hit my head on a doorframe that impact *causes* my pain. Put roughly, pain is the effect of my head hitting the doorframe. It would, however, be misleading to say that pain is *constituted* by its being an effect. Pain *is*, essentially, a certain kind of phenomenal experience that, in beings like us, is *caused* by neural pathways that can be activated by hitting our heads on doors, among other things. All this highlights a straightforward point—the *causal* relation is not the same as the *constitution* relation. To ask questions about the former is *not*, necessarily, to ask questions about the latter. Furthermore, recognizing this distinction emphasizes something that the non-causalist already admits, namely that the causal relation is *extrinsic* not *intrinsic*. But if this is true then premise (3) is on weak ground. Goetz’s claim that the *effects* of efficient causation are passive is ambiguous. Its plausibility trades on the true claim that an effect is

always passive with respect to its cause. That claim, however, concerns a property of the *extrinsic* relation between an effect and a cause. Thus, while it is absurd for an entity to possess two ‘contradictory’ essential intrinsic properties, it is possible for an entity to, at the same time, possess an intrinsic property and an extrinsic property that are in opposition.

I think Goetz would respond to this by pointing out that all of those mental events that we *know* are caused are also intrinsically passive, like our believing or perceiving. Hence, we have reason to conclude that it is *likely*, given our evidence, that if something is caused it is also essentially intrinsically passive. I have framed this as a probabilistic claim to avoid presenting Goetz as affirming the consequent. But I note that in his work he seems to present this conclusion with more certainty than is warranted. Regardless, it is manifestly clear that the *essentially* intrinsically passive nature of such *caused* mental events would have to be a necessary *a posteriori* discovery *not* a conceptual truth, as he originally claimed.

If so, then it seems we have a potential response to Goetz and any strong non-causalist who benefits from similar considerations. It is false that if a mental event is *caused* it is *essentially* intrinsically passive. It might be necessarily extrinsically passive with respect to its cause, just as *all* effects are. But it is conceivable for a necessarily extrinsically passive property to be essentially *intrinsically* active. Thus, I concur with Palmer and Ginet that the strong non-causalists have yet to give us a compelling positive argument for holding that position.

2. Dialectical Stalemates and the Incomprehensibility Challenge

Above, I considered three important causalist objections to the non-causal views I introduced in chapter 1—(i) the randomness objection, (ii) the rational explanation objection, and (iii.) the origination objection. I showed both how non-causalists respond to these objections and where those responses are weak or unavailable. Finally, in the previous section, I examined the

debates between weak non-causalists and strong non-causalists; my conclusion was that the best argument *for* strong non-causalism fails and thus there is little reason to hold such a difficult position. I now turn to present what I take to be the lesson learned from all these considerations, namely, that both strong and weak non-causalism have difficulties that render them undesirable positions, despite their effective criticism of causalist rivals. This conclusion will leave space for a new non-causal argument, that takes the strengths of Ginet's weak non-causalism but avoids the general difficulties that non-causalism faces.

Ginet's weak non-causalism failed to address the linkage objection and raised concerns about unparsimonious radically disjunctive explanations. Strong non-causalists avoid these problems by holding a more radical view. They deny the very grounds that allow for the causalist criticisms of Ginet. After all, both the linkage objection and the unparsimonious concern gain purchase on Ginet *because* he agrees that unfree actions can be caused. The strong non-causalist denies this, and instead claims that *all* actions are uncaused.⁴² This, however, requires the strong non-causalists to establish that the burden of proof is on the causalist. This is the spirit of the causalists' origination objection. The origination objection challenges the strong non-causalists to explain how it can be the case that such uncaused events appear in the world if there is no *causal* explanation for their existence. That is, it claims that the burden of proof is on the strong non-causalist. Though this objection primarily targets the strong non-causal theories, it is worth noting that since Ginet thinks that *free* action must be uncaused, the origination objection would still apply to his account of free action.

⁴² Many strong non-causalists also claim that, because all our choices are uncaused, many, and perhaps all, of our voluntary actions are free. This broadminded libertarianism, if you will, is a distinctive feature of both McCann and Goetz's views, however, we should be careful not to state it too strongly. For example, Goetz still thinks that actions can be determined, however, it is on non-causal teleological grounds that they are determined. In general, I have avoided discussing the different libertarian views of the non-causalists since this needlessly complicates my central project of examining the nature of action *as such*.

As I showed above, there are basically two positive strategies pursued by the strong non-causalists at this point. First, they can motivate a sharp dichotomy between our “practical” and “theoretical” perspective. Since, by their lights, the practical perspective is not explainable in causal terms and since we are very committed to the practical perspective, we are thus justified in believing that our actions are uncaused. This is the strategy that McCann utilizes. Second, they can try to argue from independent reasons that we should accept that our actions have a unique uncaused character as a matter of their essence. Hence, any theory that attempts to do justice to the phenomenon human action would *have* to be a non-causal theory. This is the strategy that Goetz deploys.

Again, as argued above, these strong non-causalist arguments do not adequately address the burden shifting requirement. I focused on Goetz’s attempt to ground this argument in a certain ontology of mental powers and how this was supposed to provide evidence that our actions *must* be uncaused. This argument failed to provide the support that Goetz proposed. The strong non-causalist could, however, dig in their heels at this point and maintain that their positions are not unreasonable. Even if they do not have a full non-question begging argument for their starting assumptions that does not, by itself, undermine the viability of their theory. Indeed, at various places both McCann and Goetz do exactly this, adverting to *tu quoque* responses as well as claiming their opponents “beg the question.”

I noted these responses with an unfavorable tone throughout the above discussion, but it is important to clarify *why* I believe such responses are inadequate. This gets to the central problem with *many* contemporary non-causal responses and how they fit into the overall causalist/non-causalist discussion. I contend that such responses contribute to a “dialectical stalemate” that undermines the development of the action theory debate as a whole.

2.1 Dialectical Stalemates

The term “dialectical stalemate” was made popular by Fischer (1994; 2006), who used it in the context of discussing the free will debate. Fischer (2006) says a dialectical stalemate is a situation in which some argument or family of arguments cannot be given further support without begging the question against an opponent of that argument. Any *decisive* evidence either for or against the argument would presume the very thing under disagreement.⁴³ Fischer concludes that, “one of the most salient characteristics of a perennial philosophical problem is that it involves a Dialectical Stalemate. Further, it seems that the free will problem is a true philosophical classic in part because it is an environment rich with in Dialectical Stalemates” (p. 84).⁴⁴

The philosophical debates concerning the nature of action and causation’s relation to it are a good example of just such a “perennial philosophical problem.” Furthermore, the strong non-causalist strategies described above are examples of the kind of strategies that lead to dialectical stalemates. After all, Goetz explicitly admits that there are no strategies for defending

⁴³ Fischer considers, as an example, the claim—implicit in Van Inwagen’s famous consequence argument—that if both the past and the laws of nature are not up to us then if they jointly entail a unique future that future is *also* not up to us. This is an instance of a more general principle called the *Transfer of Powerlessness* principle, which states, roughly, that if X is not up to us, and X entails Y, and X’s entailing Y is not up to us, then Y is also not up to us. Fischer imagines that someone might attempt to decisively undermine the principle by pointing to an ordinary context, like S mows the lawn at t_2 and determinism obtains. The opponent to Transfer then alleges that it is *obvious* that S can at t_2 refrain from mowing the lawn. Though this example, if plausible, provides a decisive reason to reject Transfer it does so by begging the question against the incompatibilist. One cannot simply assume that although determinism obtains S is able to refrain from mowing the lawn at t_2 no matter how *obvious* the proposed example is supposed to be.

⁴⁴ There is a related dialectical impasse that is not vicious, though perhaps more frustrating. It is sometimes summed up with the dictum “one philosopher’s *modus ponens* is another philosopher’s *modus tollens*” often attributed to Putnam (1994, p. 280). When this happens the proponent of conclusion, C, might argue. If P then C, *obviously* P, therefore C; in contrast the opponent of C might agree that If P then C, but *obviously* not-C, therefore not-P. This is not flatfootedly question begging, but it can still lead to a stagnation of the dialectic as people debate the *obviousness* of the truth of premises. Further, one might worry that at this point a true dialectical stalemate will follow. For, it is hard to see how one *could* establish the obviousness of one of those premises without begging the question for the opponent.

his view that are not question begging. The deeper question here, however, is whether this is a problem.

Goetz (2008), as a representative strong non-causalist example, seems to think that such stalemates are no great loss. As he claims, “any philosophical position must start somewhere, and in the case of views about [non-causal] freedom, that somewhere is with certain fundamental experiences of being an agent of a particular kind” (p. 5). He explicitly calls this a kind of “stalemate” but contends that the best one can do is provide the relevant defensive arguments.⁴⁵ McCann (2012) suggests a similar response. He claims that non-causalists like him take the practical perspective seriously, despite the inability of reconciling it with our best causal understanding of the natural world (pp. 255-261).

There are two reasons to consider dialectical stalemates an undesirable outcome for any philosophical debate: (i.) they undermine the ideal goal of philosophical debate and (ii.) they require us to posit an error theory for why there *is* a stalemate. First, the idealized goal of philosophical debate is to arrive at an intersubjective, stable, and reflective judgment about the phenomenon under consideration, which by its very intersubjectivity, stability and reflectiveness is most likely to be true. Put in less technical terms, the ideal aim of philosophical debate is to arrive at truth via reflective consensus building among the various parties of a philosophical dispute.⁴⁶ If this is the *ideal* then the dialectical stalemates are, as such, non-ideal and bad for that

⁴⁵ This penchant for burden shifting is even more clear in his discussion of Davidson style objections to non-causalism, in particular those offered by Mele. In that discussion Goetz (2008) says, “before proceeding further, it is important to make clear that my treatment of Davidson’s argument is strictly defensive in nature. In other words, it is not my point to prove or argue for the view that the reasons that explain choices cannot be causes of those choices. I cannot provide such a proof, but this inability is not relevant, after all, it is Davidson who has claimed that unless one accepts a causal understanding of reasons, one cannot account for the distinction between having a reason and choosing with it and having a reason and choosing because of it” (p. 45).

⁴⁶ These two sentences make *enormous* claims regarding the nature and purpose of philosophical methodology that I recognize are controversial. I am without space to argue in detail on behalf of this approach, as such I hope their intuitive plausibility as at least *one* viable articulation of the ideal goals of philosophical argument is useful. I frame philosophical method in this way by drawing on Rawls’ reflective equilibrium methodology as well as his claim that

reason. It is possible, however, that this badness is merely a feature of some intrinsic limitation in the subject matter of this *particular* debate, rendering it non-ideal in a less pernicious way. For example, it is in some sense non-ideal that we cannot ensure that we *never* make a mistake in our reasoning, however, this badness is merely an intrinsic limitation of our status as human reasoners and thus not a matter to lose sleep over. As such, when evaluating the dialectical merits of arguments that entail stalemates it matters *why* and *how* a debate has arrived at that point.

Given the above considerations, there are at least two options: (1) fundamental experiences, which ground the starting premises between these two views in action theory, are irreconcilable because one side is fundamentally deceived. (2) both sides have a partly veridical understanding of the phenomena under discussion and because of this they are reticent to let go of their views in the face of potential counter-examples. If so, then the stalemate indicates that an alternative view that synthesizes the successful bits of each theory is to be preferred.

Both possibilities are, in a sense, bad. But, scenario (1) is *tragically* bad, because it means that the most justified response is merely to provide the best defensive arguments for your view and hope that you are not the one who is fundamentally deceived. Notice, this is analogous to the *intrinsic* limitation discussed above, since there is nothing that can be said to the agent that does not have that experience. Unlike the intrinsic limitation, however, this is still *tragic* because it is not evenly distributed across reasoners. In my intrinsic limitation example, we assume that *all* agents are non-ideal in the sense that they might make mistakes in reasoning. In contrast, scenario (1) claims that only *some* agents are fundamentally deceived.

political philosophy can be understood as a kind of “reconciliation.” In addition, I was motivated by Lloyd’s discussion of the relationship between intersubjectivity and objectivity in science. For readers interested in exploring these methodological concerns further I recommend Lloyd (1995) and Rawls (1951; 1971, pp. 19-22, p. 580; 2001, pp. 1-5) as a starting place. To repeat what I said in the Introduction, I am not sure this is the *only* appropriate philosophical methodology. Indeed, since this is an idealized conception, certain non-ideal circumstances might require different methods. Furthermore, there might be debates that involve philosophy that are not themselves “philosophical debates” in the technical sense and thus require different methods.

If this is indeed the proper explanation for why the debate is in a dialectical stalemate then a response like Goetz would be justified. The stalemate would be intractable and thus, though non-ideal, the best we could do is provide defensive arguments and hope we are not the deceived reasoners. Again, this is analogous to the intrinsic limitation problem discussed above since it makes the dialectical stalemate a necessary feature of the debate and therefore to be accommodated rather than resolved, however, it differs with respect to its badness—it is tragically bad.

In contrast, scenario (2) is tractable. The difficulty can be resolved precisely because the stalemate results in a certain philosophical dead-end that is psychologically difficult for the proponents of the various views to exit. On scenario (2) what is required is not resignation but philosophical creativity. Fischer seems to think *most* instances of dialectical stalemate are like scenario (2); however, I agree with Goetz that scenarios like (1) are possible. Given this, we must ask, what does this tragic badness of scenario (1) consist in? How is it different from mere non-ideal limitation? By answering these questions, we can begin to determine if we have reached a point in the given dialectic wherein we are *justified* in accepting the stalemate as inevitable.

Roughly, scenario (1) is tragically bad because it requires doubt in two directions. First, it requires that one fundamentally doubt the status of your opponent as a knower. You must assume they are fundamentally deceived and thus the philosophical ideal of intersubjective, reflective, consensus is unavailable *as such*. Furthermore, it requires you doubt your own view. By this I do not mean the normal doubt that you might be in error and must be corrected. Instead, you must consider that you are *fundamentally deceived*. Again, like your philosophical opponent your status as a knower might be fundamentally flawed. Thus, *if* we are in this situation with respect

to a dialect stalemate then the very possibility of arriving at truth via *argument*—since as I claimed earlier truth *just is* arrived at via stable, intersubjective, and reflective, judgement—is undermined. Put differently, if our view is, in fact, true we come to its truth (in some sense) accidentally and we justifiably hold it to be true only because all views so situated, *even contradictory views*, could also be justifiably held.⁴⁷

This state is undesirable and, as such, it puts *pressure* on any claim that the dialectic is intractable *because* of such a fundamental difference in starting premises. If one can achieve the desideratum of your position *without* contributing to such an intractable stalemate then it is both reasonable to prefer this outcome and, furthermore, the norms of philosophical dialectic require that one pursue such a theory.

This, then, is the *first* reason why the strong non-causalist contribution to dialectic stalemate should motivate us to seek an alternative view that still captures what is appealing about non-causalism but does not undermine the philosophical dialectic itself. There is, however, a more mundane reason. The strong non-causalist view requires establishing an error theory for why their initial premises, that the causalist beg, are not as self-evident for their causalist opponents. All the strong non-causalist offer something like this, with McCann’s attempt being the most nuanced. Error theories, however, are both difficult to produce and are generally thought to be justified only as a last resort. Furthermore, one might deploy a “parsimonious” objection. If one could explain why we’ve arrived at a dialectical stalemate and then resolve the stalemate all without adverting to an error theory this solution would be simpler than crafting both a theory *and* an error theory for why your opponent does not share your initial premises.

⁴⁷ This is assuming again that we are in the situation described by Goetz where all contradictory entailments are avoided or resolved and all that is at stake is a fundamental premise that cannot be argued for.

To summarize, the strong non-causalist response is undesirable because the *sort of* dialectical stalemate it entails both makes its philosophical defense unparsimonious *and*, more seriously, because it undermines the ideal of philosophical dialectic as such. Of course, as noted above, it is *possible* that this undesirable position is, in fact, the correct one and our situation is just tragic. However, the above considerations should *motivate us* to only accept this conclusion after we've attempted to (a) provide an alternative account of the dialectical stalemate and (b) provide an alternative synthetic theory that attempts to resolve the stalemate without losing what is desirable in current theories.

My task, then, is to establish (a) and (b). In the remainder of this chapter I argue for (a) by showing that the dialectical stalemate between causalists and non-causalists is endemic to current non-causal theories because of their reliance on experiential premises that necessarily obscure and thus plant a seed of incomprehensibility that can only be addressed by adverting to the tragic stalemate described above. This incomprehensibility objection identifies the *ground of* dialectical stalemate and thus provides a clue to how this situation might be resolved. Furthermore, Ginet stands as an outlier among the non-causalists I consider, insofar as his weak non-causalism seems to avoid the central thrust of the incomprehensibility objection. I argue, however, that this actually highlights why the objection works against other non-causal theories and in turn reveals that Ginet's own attempt has similar flaws that avoid the objection only accidentally. Having established (a) I then turn in the next chapter—chapter 3—to the project of (b), establishing an alternative account that avoids the dialectical stalemate while simultaneously maintaining the strengths of the current accounts.

2.2 The Incomprehensibility Challenge

Why, then, do these strong non-causal views lead to dialectical stalemates? I contend, first, that all the proceeding strong non-causals accounts share an argumentative structure that is crucial for their positive accounts. They all depend on, what I call, *experiential arguments* to support the conclusion that our simple mental acts, or decisions, or choices, are uncaused.

Consider again, for example, the following passages from McCann, and Goetz:

It is *self-contradictory* for me to assert that I inadvertently or accidentally decide anything, or to pretend that I am ever passive in deciding . . . The positive marks of this are two. First, deciding has in our phenomenal experience an inherent spontaneity. When we decide we feel that we are doing rather than being done to, that we are being active rather than passive. Second, deciding is intrinsically intentional: when we decide we *mean* to be deciding, and we mean to decide exactly as we do. (McCann 2012, p. 254.)

[A choice] is simply immediately apprehended or experienced by its agent as intrinsically active in nature and, thereby, as uncaused, because what is intrinsically active is essentially uncaused . . . Though a choice, unlike an experience of pain, lacks any qualitative feel, I think the view that it too has a nonrelational, intrinsically active nature that is introspectively accessible to its agent is bolstered by an epistemological datum of mental action . . . it is an epistemological feature of an agent who knows that he is making a choice that he knows this while he is choosing . . . if he knows in this way that he is choosing, then it seems to follow that choosing is intrinsically different from a mere happening or passive event and that he is *aware* of this difference. (Goetz 2008, pp.11-12)

These passages offer key defenses of non-causal positions against a causal skepticism. In them, the authors appeal to experiential data as crucial evidence for their non-causal account. Our simple mental actions are experienced as *intrinsically active* and this provides evidence that any causal account must be inadequate.

From this general frame we can see two important claims affirmed by all strong non-causalists. First, the nature of a mental event is introspectively accessible to us. This claim, though bold, is in line with other common claims in philosophy of mind, particularly in the discussions of *qualia*. Of course, simple mental acts do not have *qualia* in this same sense of

pain but I take the non-causalist to be making a similar point. Simple mental acts are *known* or are *referred* to by reference to our experience of them *in acting* and any theory that attempts to abstract away from this detail is, in the words of Frankfurt “[directing] attention exclusively away from the events whose natures are at issue, and away from the times at which they occur.” (Frankfurt, cited in Goetz). It is *constitutive* of simple mental acts that they are *active* and this *active nature* must be considered.

Second, the strong non-causalists rule out, indiscriminately, *all* extrinsic relationships as a possible explanation for the essence of simple mental acts. Though the target of their arguments is causation as understood by causalists, even non-causal relations, if extrinsic, are unsuitable candidates for explaining simple mental acts by the lights of these considerations. Furthermore, the non-causalists have implicitly suggested that there are no *other* intrinsic features that could define these acts other than their active nature. Hence, theorist like McCann are left with the experiential features to wholly define the nature of such actions, whereas Goetz tries to say a bit more by positing it as a conceptual truth that exercising a mental power entails such active experiences.

Before examining how this contributes to the dialectical stalemate, it is important to note that Ginet makes similar claims. Though Ginet’s weak non-causalism allows for extrinsic causal relations to explain the existence of our simple mental acts, he too believes that such relations are not sufficient for explaining the nature of action. Indeed, his claims regarding what *makes* an event an action similarly appeal to intrinsic and experiential features of agency (consider his extended discussion of mentally saying the French word *peu* rather than having it come unbidden to your mind). Again, as he states:

The mental act has what we may call (for lack of a better term) an *actish* phenomenal quality. This is an extremely familiar quality, recognizable in all mental

actions . . . This quality is intrinsic to and inseparable from the occurrence of the word in my mind when I mentally say it. It belongs to the manner in which the word occurs in my mind and is not a distinct phenomenon that precedes or accompanies the occurrence of the word. (Ginet, 1990 p. 13)

This means that even though there is an important difference between Ginet's views and those of the strong non-causalists there is a commonality as well, one that I consider again below.

First, returning to the strong non-causalists, I claim that their reliance on experiential premises contributes to the dialectical stalemate between causalist and non-causalist theories. If an approach *depends* on experiential datum that is disputable, there is a sense in which such a view is *required* to move to a "begging the question" style response. If one of your premises is not, really, open for debate this pushes the dialectic into a position where the real point of contention is whether the relevant experiential datum *is* what is experienced or how such datum might be illusory. From this argumentative dead-end there is no easy way out—a genuine dialectical stalemate.

McCann (2012) articulates this worry well by recognizing a fundamental divide between the, so-called, "theoretical perspective" and the "practical perspective" exemplified in the non-causalism debate (pp. 251-253). He admits, it is unclear how experientially grounded views could fit with our causally informed understanding of the natural world in a manner that was explanatorily satisfying for someone who didn't share the non-causalists' experience. I call this gap in explanation the *incomprehensibility challenge*. It is important to note that this is not a *counter-example* nor is it a strong objection in the logical sense. Rather, the incomprehensibility challenge is a worry about explanatory mysteriousness and *how* this mysteriousness engenders the philosophical dead-end of dialectical stalemates. My claim is that the strong non-causalists' very *retreat* into an unarguable "begging the question" style defense is made rational by the experiential heart of their arguments—how else might you defend one's experience if it diverges

from your opponents? Thus, the road to dialectical stalemate is, though not entailed, at least endemic to strong non-causal views.

I return now to the weak non-causalist position. As I've shown above, the strong non-causalists arguments against the possibility of some actions being caused are ineffectual because they support their key and most controversial premises by Kripkean style reasoning that doesn't hold-up under scrutiny. In responses to this the strong non-causalists can still retreat into purely defensive arguments that accuse their opponent of begging the question. I also argued above that this strategy is undesirable because of how it contributes to dialectical stalemates, furthermore, these stalemates are an endemic feature of the strong noncausalist argumentative strategy because they rely on experiential datum to ground key claims regarding action—this last claim is what I call the incomprehensibility challenge. With this in mind, how does weak non-causalism fair?

I follow Ginet and Palmer in thinking that the weak non-causalist position is defensible against many standard causalist objections. I do, however, think that Ginet's version does not have an adequate response *either* to Clarke's version of the linkage objection *or* worries about parsimony. At first glance, it seems that Ginet at least avoids the charge of dialectical stalemate that I raised before. Once one examines many of the standard criticisms of Ginet's view, however, it becomes clear that underlying worries about the obscurity of experiential premises remain. These kinds of criticisms come from two directions. First, as in the criticisms examined above, Ginet's opponents sometimes focus on how the phenomenal quality he describes is *too thin* to provide an adequate explanation for what *makes* an event into an action. Second, insofar as *free* action is uncaused on his view, opponents can raise similar concerns regarding the

obscurity of phenomenal premises and how they might explain the production of actions. In short, the reliance on experiential evidence generates much skepticism, even in Ginet.

As O'Connor (2000) notes, "[Ginet's style of non-causalism] is implausible, in much the same way that phenomenalism about physical reality is implausible: seemings are not sufficient for realities." (p. 26). Lurking behind this claim is a concern that non-causal accounts are either empty or unnatural. For example, though Ginet (1990) contends that non-causalists use productive language (e.g. 'doing,' 'making' etc.) in a radically qualified manner (pp. 13-14), causalists, like O'Connor (2000) claim that such non-causal explanations are merely parasitic on implicit causal notions (pp. 25-26). Pereboom (2014) makes a similar point (pp. 39-43).

Further, skeptics claim uncaused events cannot fit into our causally structured understanding of the natural world. O'Connor (2000) claims such uncaused events would, given the assumption that fundamental physical processes are causally connected, be inconsistent with both materialist and emergent dualist accounts (p. 27). Presumably because of how it intruded in the "naturalistic" causal chain. Kane (1996) makes a similar claim stating that, in the case of free action, causal explanations must be part of the story for actions to fit into the natural order (p. 174). Ginet (2014) responds to these concerns by pointing out there is no *a priori* reason to hold that kind of skepticism (p. 25). While I agree with Ginet's sentiment, I note that his response comes close to the kind of "question begging" defensive strategy pursued by the strong non-causalists. As noted above, this defensive strategy is undesirable, all other things equal, because of how it contributes to dialectical stalemates.

The fact that Ginet's views are not *as obscure* as the strong non-causalists does point us towards a potential solution. Ginet's weak non-causalism allows for the *possibility* of extrinsic relations—in this case causal relations—for bringing about actions. This means that, minimally,

there is no explanatory mystery regarding how an (unfree) action is brought about, even if what *makes* it an action appeals to something experiential. This explanatory power partly explains why the causalist responses to Ginet *must* be more nuanced (e.g. Clarke, Mele, etc.). It becomes obscure precisely where Ginet starts to rely on experiential datum to do philosophically explanatory work regarding the *nature* of action and *production* of free action. Hence, if a non-causalists wants to avoid the charge of incomprehensibility they should avoid *relying* on experiential datum as the primary explanatory feature and consider the possibility of non-causal *extrinsic* relations.

The experiential premises used in non-causal theories of action are obscure and merely relying on contrary intuitions is unlikely to persuade many philosophers. Since it is in actions that agents interact with our causally structured reality explaining actions with something incomprehensible from this theoretical perspective is especially unsatisfying. So, given the incomprehensibility challenge's force, it is important for non-causalists to craft a more plausible positive account. This establishes (a), which—as I mentioned above—is an alternative account of dialectical stalemates that shows how non-causal theories contribute to their existence. Now the ground is clear to establish (b), which is a theory that resolves the dialectical stalemate without losing what is valuable in current theories. In chapter 3, I argue for such a view at length. I propose that non-causal theories ought to embrace extrinsic non-causal metaphysical relations, like essential metaphysical dependence, to explain how uncaused actions can both belong to the agent *and* be part of our causally structured understanding of the world.

CHAPTER 3: ACTION AS ESSENTIAL METAPHYSICAL DEPENDENCE

Let me briefly take stock of where the argument stands before turning to my essential metaphysical dependence account of action. What makes the event of my walking around the block an action of mine rather than merely an event I am the subject of or, more impersonally, a mere happening in the world? I distinguish this from the related question, “what does it mean to act for reasons?” The first sort of question is about defining what an action *is* and the second sort is about understanding what it is to *act for a reason*. In many accounts of action these questions collapse together. For example, on many standard views reasons, conceived as causes, both provide action explanation (e.g. answer the question, “why did you do x?”) *and* reveal what makes an event count as an action (e.g. answer the question, “what defines x?”).⁴⁸ A further question regarding action is, “what are the necessary and sufficient conditions for x?”, this is a broader question than the definitional question though they are related.

In this chapter, my *primary* goal is to provide a unique non-causal answer to the question “what makes an event count as an action?” First, in §1, I review the traditional accounts of action from chapter 1 and chapter 2 and my reasons for rejecting them. In §2, I present my initial account of action as essential metaphysical dependence and clarify the mental ontology that my account presupposes. I then turn, in §3, to explain key concepts used in my definition, including dependence, essence, grounding, and their relationship with causation. Once these things are clarified, I articulate in §4 the final version of my account. In §5 I argue first that my account successfully meets non-causal desideratum and then defend it against causalist objection. Finally, in §6 and §7 I consider three significant criticisms of my account—the circularity objection, the

⁴⁸ I note that the sense of definition I am appealing to here is not merely semantic or conceptual. I follow Fine in thinking that there is a more metaphysically laden sense of definition (owing to Aristotle) that involves coming to terms with what an entities essence or identity might be.

control objection, and doubts about its compatibility with causation—and provide arguments to show that they are unsuccessful.

1. A Brief Recap

What sorts of events are the appropriate target for defining actions? As noted in chapter 1, mental events, which always underlie our overt physical actions, are the basic events that we are attempting to understand. As argued in previous chapters, there are four different ways to answer the definitional question regarding action:

- (1) **Extrinsic Causal Accounts:** Event, e , is an action of an agent, S , iff e is caused (in an appropriate way) by some other appropriate state, s_I (e.g. desire and belief, or intention).
- (2) **Intrinsic Event-mediated Causal Accounts:** Event, e , is an action iff S is the subject of an agent-involving event, e_I , and e_I causes e , such that e consists in S causing something via e_I .
- (3) **Intrinsic Agential Causal Accounts:** Event, e , is an action iff S directly causes e as a substance rather than as the subject of an event.
- (4) **Non-causal Accounts:** Event, e , is an action iff e has an intrinsic non-causal feature or arises from an appropriate non-causal power. (e.g. e has an actish phenomenal quality; e is a non-causal result of an agent's active power to choose, etc.)

I pause to re-emphasize that the difference between extrinsic causal accounts (e.g. 1) and intrinsic causal accounts (e.g. 2, and 3) lies in *how* the causal relation explains why an event counts as an action. Extrinsic theories say that an event is an action because it is *caused* by S in an appropriate way; in contrast, intrinsic causal theories claim that an event is an action because it *consists* of S 's causing something. Thus, extrinsic accounts are so named because what makes an event an action is something *external* to the act—namely, how it was caused. In contrast, intrinsic accounts are so named because what makes an event an action is *internal* to the act—the action itself consists in S causing something. In my discussion, in chapters 1 and 2, I considered reasons to reject all of these proposals.

Again, to briefly recap: definitions like (1) are subject to counter-examples, in part because of how difficult it is to specify the “appropriate” causal relation between the desire and belief or intention and the mental event it is supposed to make into an action.

On proposal (2) the event described in this analysis is described *either* as a desire-belief pair *or* as a neural event. If it is a desire-belief pair then familiar counter-examples regarding acting without the relevant desire or belief or intention arise.⁴⁹ If it is a neural-event then there is a dilemma. On the one hand, since neural events are opaque to the agent, they would be unjustified in claiming they acted since they have no knowledge of the event that makes a mental event into an action. On the other hand, claiming that neural events are *identical with* occurrent mental events that initiate action leads to a regress. In attempting to define action, we would have merely said that mental events are actions just in case they are caused by agent-involving neural events that are *also* actions (for, to try to act is to act, as Ginet points out). Either way, the proposal fails.

Definition (3) does better by ascribing to the *agent* a direct causal link that then makes the mental event, *e*, into an action. These agent-causal views, however, are inadequate for several reasons. First, it is unclear how agent-causal views can successfully respond to the explanatory challenges raised by Broad, among others. More directly, the obtaining of an agent-causal relation is *not* explainable in causal terms on pain of a regress. But, if the relation stops at some point then it seems the ultimate explanation for agential control is non-causal, the very thing the agent causalist is attempting to avoid.

Definition (4) is, I believe, on the right track. What makes a mental event an action must be non-causal. There are two reasons for this: first, because the distinctive practical character of

⁴⁹ Such counter-examples, I note, would also undermine proposal (1).

mental acts does not imply or require causal explanations and instead is in tension with such explanations; second, because the counter-examples to causal accounts of what makes a mental event into an action are compelling and wide ranging.

Current non-causal proposals, however, are inadequate. In chapter 2 I argued that non-causal views assume a dialectical position that typifies what Fischer (1994; 2006) calls a “dialectical stalemate.” This problem is worse for strong non-causalism. Strong non-causalists assume that the perspective of agency is unreconcilable to our causally informed understandings of the natural world, thus, criticisms of their view amount to mere question begging. Goetz admits this strategy, noting openly that all his defensive arguments involve either accusing his opponent of begging the question or a *tu quoque* response. These sorts of responses are unsatisfying and causalists are rightly suspicious that the non-causalists are merely avoiding and not addressing their criticisms.

In contrast, weak non-causalism attempts to find common ground with the causalist by conceding that the non-causal features that define action are compatible with the “causal nexus” of nature. Such weak non-causal views, however, are undermined by arguments that show non-causalist explanations are superfluous. For example, in chapter 2 I argue that Ginet’s account does not adequately address criticisms raised by Clarke regarding the explanatory connection between intentions and actions. Moreover, even when discussing free action Ginet’s attempts to articulate a response to the “unnaturalness” objection is at best a sketch. Still, his sketch is suggestive, as he says:

“Nature has held many surprises. I see no a priori reason for thinking it impossible that one surprise nature might hold in store for us is that some events are not caused, that empirical investigation will persuade us that the best account of the laws of nature dictates for some kinds of events that their antecedents do not *causally produce* them but only limit what the ensuing event will be to one or another of a certain proper subset of these metaphysical possibilities.” (p. 25)

My argument fills the gap Ginet gestures towards. I explain, more adequately, what these “limiting” conditions might be and how they interact with the causal structure of the world. I argue these conditions are best conceived of as metaphysical dependence relations and they provide the best explanation for what makes mental events into actions. Moreover, I argue that the phenomenal or experiential evidence utilized (in different ways) by McCann, Goetz, and Ginet provides *evidence* for this underlying dependence relation and is not, by itself, sufficient for defining action.

2. Action as Essential Metaphysical Dependence

My view is motivated by two primary concerns. First, there is something correct about the agential focus of agent-causal views. Regardless of the inadequacy of agent-causal accounts, tying action directly to the agent is important for both explaining the nature of action *and* maintaining morally substantive agency. Second, the non-causal arguments against wholly causal explanations of action are insightful and offer the best explanation for the distinctive character of our practical lives. My theory provides a definition of action that unites these two insights. I begin by articulating my essential dependence account in rough form. I then explain in detail the components of my account (e.g. metaphysical dependence relations, interactive mental properties, etc.). Finally, I consider potential objections.

Minds have mental properties. These properties are correlated with functions of the mind. For example, plausibly, minds have properties that give rise to mental processes that involve *responding* to the world, such as belief formation, perception, and desires. In like manner, minds have properties that govern mental processes that involve *interacting* with the world, such as intention formation, willing, and deliberation. These two *kinds* of properties, “responsive properties” and “interactive properties,” roughly track the distinction that Goetz discusses

between “capacities” and “powers,” I avoid using that terminology for two reasons. First, these words *connote* more than is warranted. In particular, the language of “powers” has a distinct “causal” flavor. To avoid such connotations, I speak of these properties in terms of whether they respond or interact with the world while avoiding any suggestions about the *nature* of these connections.⁵⁰ Second, Goetz suggests these are exhaustive categories for mental activity. I am not so confident that I have access, introspectively or otherwise, to the nature of all of the mental properties that exist in minds. As such, my claim regarding mental ontology is rather limited: (i.) there are mental properties (ii.) there are at least two types of mental properties that are differentiated by how they relate with the external world: responsive properties and interactive properties.⁵¹

Given this mental ontology, I return to the central question: what makes a mental event into an action? My account, roughly, is as follows:

Action as Essential Metaphysical Dependence (version 1.): A mental event *e* is an action iff *e* essentially metaphysically depends on interactive mental properties, *P*.

This means that when we have an intention, for example, there is an associated mental event “deciding” by which we formed that intention, which counts as an action of ours by virtue of essentially depending on the interactive property (or interactive mental properties) possessed by minds. In like manner, my following an action plan or deciding a course of action has an associated mental event, a “volition”, that is an action of mine by virtue of essentially depending on an interactive property possessed by minds.

⁵⁰ One might worry that “interaction” also has causal connotations. I do not share this intuition, it seems to me that we have many examples of non-causal interactions in pure or abstract domains (i.e. mathematical or geometric relationships; logical connections, etc.), however, it suffices for the reader to recognize that by “interact” I do not necessarily mean something *causal* I merely mean to point out *that* there is a relationship of some sort between the properties involved and the world.

⁵¹ A brief word of metaphysical caution: we should not allow our ability to conceptually distinguish between “responsive properties” and “interactive properties” to give the false impression that these mental properties are separable or otherwise deeply divided faculties. Clearly these properties work together in complex ways.

I described these the interactive mental property by pointing to the functional role it plays in interacting with the world. One tempting way to characterize this entire category is as the *reasons-responsive mechanism* of minds. That is, to think of the “interactive mental property” as the property that contributes to (or perhaps constitutes) the reasons-responsive mechanism of minds. This terminology is somewhat unfortunate since “response” connotes an overly passive stance towards the reasons that an agent is confronted with. There is a deeper reason, however, for resisting this characterization. As I argued in chapter 2, reducing agency to some sort of underlying mechanism leaves theories of action unable to adequately explain actions. Given this, though there are important overlaps between my conception of “interactive mental properties” and what some compatibilists call the “reasons-responsive mechanism,” I do not believe it is helpful to characterize the properties in those terms.⁵² Instead, the properties I am discussing are best understood as those properties that constitute our agency and they should therefore be treated non-reductively.

Given my view, a natural criticism is that my talk of “interactive properties” is armchair psychology at its worst. I should be able to *name* the kind of properties I am talking about, says the skeptic, and if I cannot then such metaphysical extravagances are unwarranted. After all, I am not a neuroscientist discovering features of the mind and, as such, positing something’s existence would be *ad hoc*.

I concur that defining into existence novel features of the mind would be *ad hoc*. Hence, I claim that these “interactive properties” are merely a specification of mental elements to which

⁵² McKenna (2013) raises a similar criticism against compatibilist views, like Fischer and Ravizza’s (1999), which appeal to “reason-responsive mechanisms.” McKenna’s (2013) objection is that a mechanism-based account of reasons-responsiveness requires a principled basis for mechanism individuation, but the viability of such a project is doubtful given the level of plasticity the mechanism needs in these accounts (p. 170). McKenna goes on to propose his own “agent-based” reasons-responsive theory, with which my “agency-first” account shares some similarities.

we are already committed. I explain this, first, by noting some general features of properties that are in the background of my discussion. Second, I examine some *specific* features of mental properties that make my view plausible and in-line with our current understanding of minds.

Properties are entities which can be “predicated of” or “exemplified by” objects (or other entities).⁵³ So, when I say of an object that it is “*blue*” and “*spherical*” I’ve “predicated” or attributed both of these properties to that object. Furthermore, if my claims are *true* then that thing “exemplifies” those properties of “*blueness*” and “*sphericalness*”—a blue ball, for example. Properties are not events. The exemplification of a property may be an event (e.g. my coming to exemplify the property of “*redness*” is the same as the event of my being doused with red paint) and events themselves might have properties (e.g. the event of the explosion had properties like “*bombing*” or “*white fire,*” etc.), but properties themselves are not events. Indeed, many properties are going to be states or state-like. For example, once the tree exemplifies the property of “*greenness*” (an event) it will remain “*green*” (a state) until it changes to “*yellow*” (an event).

Finally, it is important to note that, in general, properties can stand in relations with each other. Consider the property of “*color*.” It can be predicated of an object (e.g. as in the sentence, “that object is ‘*colored*’”) and it can be exemplified by an object (e.g. that object is, in fact, “*colored*”). An object does not, however, exemplify the property “*color*” as such. Instead, to exemplify “*color*” an object must exemplify one of its determinates, like “*blue*.” In contemporary parlance, “*color*” is a determinable and *blue* is a determinate. *Blue* specifies *color*.

⁵³ Of course, as with most things in metaphysics, there are deep question concerning whether properties exist at all. And, if so, in what manner they exist. I do not engage with those debates here, preferring to instead take a moderately realist stance.

There are other relationships in which properties can stand. For example, the property of being “*cuboid*” entails other properties. Some mathematical or logical, like “*having 90 degree angles*” or “*regularity*”, and some physical when it’s exemplified in an object, like “*inability to roll on a flat surface*” or being “*stackable*.” I note that these further properties are related to the property of being “*cuboid*” by entailment or dependence *on* that property. This makes it different from the “*color/blue*” example discussed earlier.⁵⁴ Finally, many of these examples show that properties can be complex or simple. For example, this computer has the functional property of “*being a finite Turing Machine*”, but that property is itself complex and composed of other more fundamental properties.

So then, what of the “interactive mental property” that I posit? The first thing to note is that it is meant to relate or unify a type of mental phenomena. Namely, those mental *events* that are involved in interacting with the world. As such, we should ask are mental events like “*intending*,” “*willing*,” and “*attending*” unified by the “interactive mental property” in the same way that “*blue*,” “*red*,” and “*green*” are unified by “*color*”? Or, in contrast, is it more like how having “*90 degree angles*” and being “*stackable*” are related by an object exemplifying a “*cuboid*” structure? This is important because one way to read the above objection is that my “interactive mental property” is just an empty category that unifies the phenomenon of agential events (i.e. “*willings*”) by merely abstracting out their similarities without *explaining* them.

To use the above example, “*color*” is a generic or abstract property that *becomes* specified and can only thereby be exemplified. If someone—a Martian let’s say—is confused about “*blueness*” it is not explanatorily helpful to merely note that it is a “*color*.” If the

⁵⁴ An additional way that properties can be related to each other is through essentially dispositional properties. Dispositional properties just *are* essentially functional, causal, or logical relation to another property. For example, the property of *fragility* is just the disposition for something to shatter given certain stimulus conditions.

interactive mental property is related to simple mental events in a similar manner to the determinate/determinable relationship, then I have not really explained anything about either mental events or the interactive mental property.

I believe that the property of interactive mentality is not merely an empty abstraction. To show this, I offer an analogy to another commonly invoked mental property, which will clarify the explanatory work interactive mentality does in my account. Minds like ours have the property of consciousness. It is by virtue of this property of consciousness (or these properties of consciousness) that we are *able* to experience phenomenal properties like “*pain*”, “*the sensation of heat*”, and “*redness*”, for example. It is plausible that consciousness is an intrinsic and essential property for minds like ours.

Furthermore, more specific phenomenal properties are dependent on the property of consciousness for their exemplification.⁵⁵ Consciousness is not, however, reducible to the various and varied phenomenal properties that it enables minds like ours to experience. I contend that a conscious mind has both the essential property of consciousness *as well as* the exemplified phenomenal properties which depend on consciousness. Crucially, though these are distinct properties and though consciousness is the more metaphysically basic (e.g. pain properties *depend* on consciousness, not the other way around), *epistemically* speaking, we only have access to our own property of consciousness *by virtue* of having experienced the dependent and varied phenomenal properties. This should be, I think, unsurprising. The property of consciousness is mentally fundamental for conscious minds. As I argue below in section 6.1, when describing fundamental entities in a given domain often their essence is definable only in

⁵⁵ I recognize that people often mean different things by the term “consciousness,” for example as Block (1995) notes we might mean “phenomenal consciousness,” “access consciousness,” “self-consciousness,” etc. I do not engage with these subtleties directly, but in this context I am primarily referring to phenomenal consciousness.

terms of what they metaphysically ground since such fundamental entities cannot be decomposed into elements upon which *they* depend in that domain.⁵⁶

If this rough ontology is plausible, then we have a direct analogy to interactive mental properties. For, just as the property of consciousness is fundamental for conscious minds, so too is the interactive mental property fundamental to agential minds. We experience mental events like “*intending*,” “*willing*,” and “*attending*,” which can be re-described as our coming to exemplify certain agential mental properties. These properties are unified by virtue of being dependent on (grounded in) the agentially fundamental interactive mental property, which just is that property in virtue of which we have mental events like “*intendings*,” “*willings*,” and “*attendings*”. Again, we cannot merely collapse the interactive mental property into the specific agential events, in part because they are too varied. Just like the variety of phenomenal properties calls for a unifying and enabling explanation, so too do the varieties of agential properties need a ground.⁵⁷

A few final clarifications. First, these categories are meant to pick-out mental events and relationships we are familiar with both from experience and general folk psychology, but I do not pretend this explains *how* they relate to neurobiology or brain structure. A metaphysical account of mind and agency is an attempt, to misappropriate Sellars’ evocative phrase, “to understand

⁵⁶ To be clear, such entities might be explainable in terms of cross-domain relationships. So, for example, fundamental mental properties might still be dependent on physical properties and thus are not fundamental in at least one sense.

⁵⁷ There is a very similar proposal in theories of emergent minds, particularly in the work of O’Connor and Wong (2005), which posits fundamental and general mental properties that emerge and then cause further, more specific, emergent properties. As O’Connor and Wong (2005) note, “It is plausible that there are enduring baseline mental states that partially underwrite more specific and often momentary mental states. (Underlying one’s visual awareness of a computer screen, e.g., is a more general state of conscious awareness that persists when one looks in another direction. We might plausibly conjecture that underlying our entire mental lives are certain highly general states, themselves mental in character, disposing us towards having specific sorts of mental experiences and cognitive states in suitable circumstances)” (p. 665). I concur with O’Connor and Wong and my discussion of “interactive properties” as those properties that *fundamentally* constitute our agency is, in many ways, compatible with their picture.

how ‘mental things’ in the broadest possible sense of the term ‘hang together’ in the broadest possible sense of the term.” It provides interpretation that should be compatible and potentially provable or disprovable within our best understanding of *how* things work, but it is not itself (usually) an account of the mechanics of how this is to be done. To this end, my account leaves open what underlying physical structures support these mental properties. Second, it is important to recognize, at least in humans, how interrelated the conscious and interactive properties are and avoid confusing distinguishability with separability. Indeed, the first-personal or subjective awareness we have of agential mental events like “*intending*,” “*willing*”, and “*attending*” depends on consciousness. This suggests that the property of consciousness is *epistemically* fundamental in a way that the interactive mental property is not. Third, and finally, we should precisely describe minds like ours as “conscious agential minds.” This highlights that in nature these fundamental properties might be combined in different ways. In fact, it is plausible in our own world that some very simple creatures might have agential minds but without consciousness (e.g. some worms, some arthropods, etc.). The opposite—conscious minds with no agential properties—though not clearly observed in our world are at least metaphysically possible.

This at least grounds the initial plausibility of the mental ontology upon which my account is based, however, this skeletal definition still requires clarification on several points. Most importantly, I must explain the *relationship* between these mental events. What are “dependence” relations? I consider this general question in the following section, before turning to describe essential metaphysical dependence in particular.

3. Dependence: An Overview

Notions of “dependence” are found throughout contemporary philosophical discussions regardless of the domain. To give some representative examples:⁵⁸

- **Logical/mathematical:** A non-empty set, $[x, x_2, \dots, x_n]$, depends on its members ‘ x , x_2, \dots, x_n ’. A conjunction, $a \wedge b$, depends on its conjuncts, ‘ a ’ and ‘ b ’.
- **Metaphysical:** Dispositions of an entity, x , depends on categorical features of that entity, x . A mental state, M_1 , depends on a brain state, P_1 , for its existence. An event depends on its participants for its occurrence.
- **Nomological/ Natural:** A human being depends on some of her biological organs to live. The fact that salt is water soluble depends on Coulomb’s Law for its realization.
- **Ethical/Social:** An action’s rightness depends on the aggregate amount of happiness in the absence of unhappiness it will produce. Whether a principle is just depends on whether parties would mutually agree to uphold it under non-coercive conditions.

I do not endorse any of these representative examples as true, but they show the breadth of the “dependence” idiom in contemporary philosophy. Philosophers express these dependency claims by relational phrases, such as “because”, “by virtue of”, or “makes it the case that” as in the sentence “the singleton set {Socrates} exist *because* the person Socrates exists”, or “The fact that Socrates exists *makes it the case that* the singleton set {Socrates} exists, or “The singleton set {Socrates} exists *by virtue of* the person Socrates.” Though theorists often give subtly different accounts of these claims, they are all taken to be claims about what depends on what. I focus on the metaphysical sense of dependence.

We may ask, however, whether dependence should be analyzed in terms of more familiar metaphysical notion. As Rosen (2010) observes, though questions of dependence are prevalent throughout philosophical debates, there are ways in which the idiom is not taken seriously on its own terms. He argues that the general method of argumentation treats language of dependence as serving a heuristic function. Philosophers use dependence talk to orient readers towards their preferred philosophical thesis, and then dispense with dependence talk in favor of difference

⁵⁸ This list is inspired by the examples given in Rosen (2010) and Raven (2015).

formulations when it comes time to *explain* the relation in detail. The thought is, according to Rosen (2010), that these widespread idioms are “ultimately too unclear or too confused, or perhaps simply too exotic to figure in our first-class philosophical vocabulary” (p. 110). Instead, Rosen says, philosophers often reinterpret dependence with modal concepts such as metaphysical necessity or supervenience.

Schaffer (2009) makes a similar observation and gives an historical explanation of this tendency among contemporary philosophers. According to Schaffer there is a fundamental philosophical divide between different contemporary approaches to metaphysics. By Schaffer’s lights, most metaphysicians follow the dominant Quinean view which holds metaphysics to be concerned primarily with questions of *existence* (e.g. do properties exist?, whether meanings exist?, do numbers exist? Etc.). Contrast this with a neo-Aristotelian view which focuses on *how* things exist. What are the relationships between entities? What depends on what? Given the dominance of Quinean style metaphysics in the contemporary era questions of dependence are not given priority. Thus, Schaffer speculates, though there is a resurgence of dependence style questions in contemporary metaphysics, the background methodology is unsuited to these new concepts and so people prefer to reinterpret the concept in familiar modal terms.⁵⁹

This brief historical account by Rosen and Schaffer raise three important questions regarding dependence in general. First, can dependence be analyzed in terms of another notion (e.g. modal necessity, essence, etc.)? Second, assuming dependence does stand on its own, then

⁵⁹ As Wilson (2014) notes, however, this “just so story” about how metaphysical dependence is much discussed yet undertheorized is exaggerated. Questions of dependence *have* been discussed in detail they just are often taken up in *specific* discussions regarding metaphysical relations that bare on problems in philosophy of mind, physicalism, reduction, and fundamentality rather than the generic question of dependence that is presumed in discussions of the grounding relation. I partly agree with Wilson that the discussion of grounding has been, at times, too imprecise and even historically naïve. I hold a hybrid view. I agree with the pluralists (like Wilson) that dependence relations are specified *within* the context of a particular metaphysical puzzle, however, I also agree with the unitary theorists (like Schaffer) that these *specific* relations are species of the genus called “grounding.” How I reconcile these things is discussed below.

what are the unique metaphysical and explanatory claims that characterize dependence and how are they related? Third, supposing that dependence supports unique metaphysical and explanatory claims what is their appropriate expression? That is, what is the logic of dependence?

Both Rosen and Schaffer advocate taking seriously the language of dependence and posit a novel metaphysical relation, called “grounding”, to make sense of the disparate talk of dependence. Their thought is that *dependence* is a unique notion that is not reducible to other concepts and thus should be analyzed on its own terms. Such a view of dependence is “robust” in that it takes dependence to be irreducible to non-dependence talk, in contrast to “deflationary” accounts of dependence. I too hold a robust view of dependence, however, the sort of dependence relation I apply to action theory is only a *species* of grounding and so different from the pure grounding accounts of Rosen and Schaffer. Before I turn to *essential metaphysical dependence*, I first discuss the general distinctions amongst dependence relations. I begin by considering whether or not dependence can be analyzed in terms of other metaphysical notions, I then clarify several distinctions among dependence relations, finally I explain how grounding is supposed to clarify dependence and compare it to closely related notions like essence.

Let me address the first question, “Can dependence be interpreted in terms of another concept?” The most common attempt to reduce or explain away dependence is in modal terms, and these attempts are often situated within discussions of *existence* and *essence*. These familiar metaphysical notions correspond to two families of dependence relations, *existential dependence* and *essential dependence*. According to Correia (2008), existentially dependent entities require certain conditions to be met for them to exist at all, whereas essentially dependent entities require

certain conditions to be met for them to *be the entity they are*. That is, their identity or essence depends on certain other conditions (p.1014).⁶⁰

There are, roughly, two approaches to analyzing dependence relations. First, one family of approaches privileges *modal* considerations as the best explanation of dependence. Second, a family of approaches that emphasizes the *priority* of one member of the relation over another as the best explanation of dependence. In what follows I sketch the basic logic of these two approaches, first regarding existential dependence and then regarding essential dependence.

3.1 Existential Dependence

An existential dependent entity requires, for its existence, certain other conditions to be met. The specification of these other conditions is where much of the metaphysical work resides, but first the *relation* must be understood. In its most simple form many philosophers interpret existential dependence claims modally, with them taking one of the following two forms:

- (1) $\Box(Ex \rightarrow Ey)$.
- (2) $\Box(Ex \rightarrow \exists yFy)$.

Where ‘ \Box ’ is a necessity operator tracking metaphysical necessity (rather than natural or logical), ‘ E ’ is a one-place predicate for existence, ‘ F ’ is a general term, and the standard logical operator ‘ \rightarrow ’ for material implication. Following Correia, I read (i) as ‘ x cannot exist unless y exists’ or ‘ x rigidly necessitates y ’ and (ii) as ‘ x cannot exist unless something is an F ’ or ‘ x generically necessitates an F .’⁶¹ Both of these interpretations are meant to track certain common metaphysical claims of dependence. For example, it would be proper to say that “this desk

⁶⁰ There is disagreement regarding how independent these two types of dependence might be. According to some views, the requirements of identity and the requirements for existence amount to the same thing and thus existential and essential dependence collapse together. Even on views where these two types of dependence are in principle separable there is still often overlap, with specific notions of dependence involving both essence and existence.

⁶¹ Correia uses the language of “rigidly necessitates,” even though “rigidly” can mean other things in logic. I maintain Correia’s terminology, but readers should remember that “narrowly” or “specifically” might be an equally viable term.

cannot exist unless its constituting piece of matter exists” which could be read as “this desk rigidly necessitates its constituting piece of matter”, which is all to say the desk *existentially depends* on the matter that constitutes it. In like manner, we may say that “I cannot exist unless there exists *some* carbon atoms, water molecules, potassium, etc., which is to say that I *generically necessitate* some carbon atoms, water molecules, potassium, which is all to say that I *existentially depend* on the existence of *some or other* entities with those properties. The generic relation is quite weak and applies to almost any claims of existence, but it is useful for making sense of certain global dependence claims.⁶²

As noted above, the possible interpretations thus far considered are *modal*. That is, they attempt to make sense of the existential dependence in terms of the *modal* covariance of the two entities under consideration. Many philosophers are concerned that such accounts are too weak to make sense of genuine dependence relations. For example, it is true that, necessarily, if I exist then the fact that $2 + 3 = 5$ exists. Given the above analysis, this means that my existence rigidly necessitates the fact that $2 + 3 = 5$, or that I *existentially depend* on the fact that $2 + 3 = 5$. But is that correct? These two things, me and the fact that $2 + 3 = 5$, only modally covary because $2 + 3 = 5$ is a necessary mathematical fact. It is inevitable, as a matter of logic, that any world in which I exist is also a world in which $2 + 3 = 5$, but that should not matter for questions of dependence. Put bluntly, the fact that $2 + 3 = 5$ does not have anything to do with me. Other philosophers, however, are willing to bite this bullet and maintain that “dependence” can mean as little as

⁶² As Correia (2008) notes, there are more precise interpretations of these simple existential dependence analysis which take things like time into account:

$$(3) \Box \forall t (E_t x \rightarrow E_t y)$$

[x cannot exist at a time unless y exists at that time]

$$(4) \Box \forall t (E_t x \rightarrow \exists u (u < t \wedge E_u y))$$

[x cannot exist at a time unless y existed before that time]

These formalizations are thought to shed light on how best to interpret certain temporal dependence theories, such as Kripkean necessity of origin arguments. For the purposes of this overview, however, I will set these aside.

modal covariance. I agree with this criticism that mere modal covariance, as currently described, does not count as genuine dependence.

Hence, I concur with Rosen and Schaffer that an appropriate understanding of dependence is not reducible to modal notions such as mere modal covariance. This is because the element of *priority* (i.e. that one of the members of a dependence relation has priority over the other member) is not adequately captured by modal notions. A natural response for the deflationist is to attempt to explain dependence in terms of a concept that already has priority built in and essence has been commonly thought to explain dependence. On such views, essential dependence is the primary dependence relation and, ultimately, the essences of things themselves explain dependence in general. Essence is prior. I agree that there is a close connection between essence and dependence, however, it is not the case that it can be wholly defined in essentialist terms.

In order to evaluate these claims I briefly explain how robust dependence theorist explain dependence in terms of “grounding” as well as examine the underlying logic of ground. Then, I turn to essentialist accounts and examine what the appropriate relation between dependence and ground is, finally I present my own taxonomy of dependence, which then will allow me to apply the concept to action.

3.2 Metaphysical Dependence as Grounding

According to Rosen (2010), theories of “grounding” offer the best way to clarify dependence (p.110).⁶³ He notes that dependence, understood as grounding, is sometimes glossed as “metaphysical causation” (Fine 2012, p. 40; Schaffer 2012, p. 122, 2016, p. 50; Wilson 2017,

⁶³ Arguments against conceiving of grounding as a unique, or primitive, or even useful, metaphysical relation are legion; however, since I am *mostly* using the concept of ground as a rough tool to understand human action via metaphysical dependence, I do not engage in detail with these skeptical worries.

pp.1-2) this means that grounding has properties that make it *analogous* to causation.⁶⁴ For example, Schaffer (2016a) says grounding is a vaguely *productive* relation (it “generates” or “shapes”), it is *asymmetrical* in direction, and it supports *explanatory* claims (p. 53). Further, grounding seems to be formally characterized by *strict partial ordering* (Correia and Schnieder 2012, p. 17; Schaffer 2012, p. 55; Rosen 2010, pp.115-116), much like many theories of causation. This means that its logical structure is best modeled as an irreflexive (e.g. An entity, A, cannot ground itself), asymmetric (e.g. Entities A and B cannot ground each other), and transitive (e.g. if A grounds B, and B grounds C, then A grounds C) binary relation. It is also, usually, considered non-monotonic (e.g. If A grounds B, and C is any fact compatible with A, it does not follow that B is grounded in A *and* C) and hyperintensional (e.g. If A grounds B then, even if A and B have the same intensional content, it is *still* false to say that B grounds A).

Why have analyses of grounding developed along these lines? Discussions of ground naturally start from how grounding relations are expressed in paradigmatic explanations of dependence. This is because many philosophers (e.g. Trogon 2013; Correia and Schnieder 2012; Schafer 2016, 2012, 2009; Rodriguez-Pereyra 2005, etc.) think grounding is closely connected to a certain type of explanation, perhaps by “backing” or “underwriting” such explanations. In addition, some theorists (e.g. Dasgupta, 2017, 2014; Wilsch, 2016, 2015; Fine, 2012a; Raven, 2012, etc.) think grounding *just is* a type of metaphysical explanation.⁶⁵

Following Raven (2015), I call philosophers who think that grounding is distinguishable yet

⁶⁴ The one exception is Wilson (2018) who instead argues that grounding *just is* a kind of causation, namely metaphysical causation. I do not engage with this interesting proposal in this paper, but I am sympathetic with Wilson’s arguments.

⁶⁵ For an interesting argument against collapsing grounding into metaphysical explanation see Maurin (2018). Unlike other anti-unionists, Maurin (2018) does not argue for a positive account of grounding in separatist terms, instead providing a skeptical anti-grounding argument. She claims that once it is established that metaphysical explanation and grounding are separable we have lost most of the justification for positing the grounding relation in the first place.

closely connected to explanation “separatists” and philosophers who think grounding *just is* a type of metaphysical explanation “unionists” (p. 326).

There are two types of formalized interpretations of these sorts of explanations. Expressions of ground using the ‘because’ connective (i.e. Singleton set {Socrates} exists because the person Socrates exists) are best interpreted by a sentential operator, whereas sentences that mention a relation (i.e. Socrates grounds singleton {Socrates}) are more naturally interpreted as a predicate. This distinction in interpretation marks a deep division amongst grounding theorists. Those philosophers who want “grounds” to be metaphysically neutral both emphasize the explanatory features of ground and tend to favor the sentential interpretation (e.g. Dasgupta 2017, 2014, Fine 2012a, 2012b; Clark and Liggins 2012; Correia 2010; etc.). In contrast, those philosophers who instead take up the controversial task of specifying the metaphysical relation and its *relata* favor distinguishing between the explanatory and metaphysical roles of grounding and so favor the predicate interpretation (e.g. Audi 2012; Schaffer 2012, 2009; Rosen 2010; Rodriguez-Pereyra 2005; etc.).

This divide emphasizes another role that grounding plays in philosophical discussion. Namely, the so-called “building” project of metaphysics. Grounding is a conceptual tool to help answer the questions, “what is fundamental and how are derivative entities ‘built’ out of the fundamental entities?” Hence, the two camps of grounding theorists are, roughly, grouped into those who take the *explanatory* features of grounding to be central and those who take its *metaphysical* features (i.e. its contribution to the “building project”) to be central.⁶⁶

Given this divide, grounding theorists explicate the grounding relation in importantly different ways. I do not attempt to untangle all the significant debates between these two camps.

⁶⁶ For an extended and insightful discussion of the so-called “building” project, see Bennett (2017).

It is important for my project, however, that certain features of grounding are true. As such, it is best to think of my argument as conditional upon a certain view of grounding being correct. This view takes grounding to *not* be identical with an explanatory relation but rather that it *underwrites* an explanatory relation.⁶⁷ So, just as causation (however it is conceived) underwrites causal explanations so too does grounding underwrite grounding explanations. Thus, by agreeing that grounding explanations and grounding relations are, strictly speaking, separable I am more inclined to a predicate interpretation of grounding expressions rather than the metaphysically neutral sentential operator.

I am also motivated, in part, because I cannot stay neutral with respect to the *relata* of grounding. Explanatory theorists tend to restrict the grounding *relata* to “facts” this is in part because facts are thought to be the appropriate targets of explanation. As Dasgupta (cited in Trogon 2013) notes, something like a table cannot explain anything. The *fact* that the table exists, or the fact that it is made out of wood, are the proper *explanans* of an explanation (p. 105). While I certainly do not deny that facts may be grounds, nor deny that facts about grounded entities might *back* metaphysical explanation, I wish to be more permissive. My account is, after all, an account of what makes an *event* an action. As such, I assume that grounding can stand between entities in the world, such as events and objects.

So, then, how does the notion of ground just articulated handle paradigm cases of dependence? Some we have already seen, but I briefly reiterate in more detail using an example from philosophy of mind. Assume that non-reductive physicalism is correct. Suppose that a

⁶⁷ Of course, one might believe that if metaphysical explanation is only *backed* by grounding relations, we could provide an account of such explanation without directly discussing the grounds. This claim can be further supported by analogy to scientific explanation. After all, philosophical theorizing about scientific explanation often runs independently from theorizing about whatever relation(s) *backs* scientific explanation. For an interesting proposal along these lines, see Baron and Norton (2019).

particular mental property, ‘m’, depends on some physical properties, ‘ Γ ’, where ‘ Γ ’ indicates a list of physical properties ‘ p_1, p_2, p_3, \dots ’. So, if grounding is the correct idiom for dependence we would formalize this as ‘ $\Gamma//m$,’ where following Schaffer (2012) ‘//’ is a two place predicate indicating *full* grounding. Suppose ‘m’ is the “property of being in pain” and ‘ p_1 ’ is the property of neuro-receptors firing, we could formalize this as p_1 / m , where / indicates *partial* grounding. Formally, we want to allow for the *possibility* of entities to have a plurality of grounds to be fully grounded, even if the limit case might be a singular grounding relation.⁶⁸ As such, in this case we can formally define a partial grounding relation in terms of full grounding as follows: $p_1 / m \leftrightarrow \exists \Gamma (p_1 \in \Gamma \wedge \Gamma // m)$. So then, informally we say “the property of neuro-receptors firing *partially grounds* the mental property of being in pain *just in case* there is some list of physical properties, the property of neuro-receptors firing is an element of that list, and that list *fully grounds* the mental property of being in pain.

Does this formalization capture our intuitive notions of dependence in this case? I believe so. Given the assumption of non-reductive physicalism, to say that mental property of being in pain depends on the physical property of neuro-receptors firing is to say there exists an asymmetric relationship between the neuro-receptors and the pain. The pain could not exist (or could not have the character it does) without the neuro-receptors, but crucially this sense of “could not” must be more than mere modal covariance. For, it is also true that for some sorts of mental idealism the pain could not exist (or could not have the character it does) without the neuro-receptors, but on this alternative view this is because the mental entities (or facts) *give rise* to certain physical entities (or facts). Hence, to capture the non-reductive physicalist’s proposal requires a certain asymmetric relation.

⁶⁸ Dasgupta (2014) takes this a step further and argues that we should conceive of ground as a “many-to-many” relation. That is, both side of the grounding relation might be a plurality.

This relation is also, plausibly, hyperintensional. This means, in the context of logic, sentence locations that are necessarily coextensive still cannot be substituted with each other without changing the truth-value of the sentence. So, for example, suppose that the set “{being in pain}” and the mental property “being in pain” are necessarily coextensive. It does not follow that I could preserve the truth value of the sentence “the mental property of ‘being in pain’ depends on the property of my neural-receptors firing” by substituting it with “the set {being in pain} depends on the property of my neural-receptors firing”. The second sentence is false.⁶⁹ Moreover, statements of dependence are non-monotonic. Just because the fact or state that “my neural-receptors are firing” is compatible with other facts, entities, or states of affairs does not mean that those too ground my being in pain.

This case is irreflexive as well, since the pain does not depend on itself, indeed it would be hard to make sense of the non-reductive physicalist position if this were the case. It is also, plausibly, transitive. Insofar as my being in pain makes it the case that I cannot focus on writing, it seems that my inability to focus on writing partly depends on my neuro-receptors.⁷⁰ Also, that my being in pain depends on my neuro-receptors firing does not exclude the possibility that there are other physical features on which it depends (e.g. certain neural-properties that ground consciousness more generally). Finally, all of the above features, especially the hyperintensionality and non-monotonicity, *back* a certain kind of explanation. By citing my neural-receptors firing I have *explained* something about my “being in pain”, something about how it depends on other features of the world.

⁶⁹ Though this hyperintensionality is primarily an expression in logic, there may be reasons to think this is tracking a real distinction as well. Nolan (2014) recently argued that hyperintensional developments underwrite a “hyperintensional metaphysics,” which mirrors the way intensional developments in logic underwrote the revolutionary developments in possible world metaphysics during the second half of the 20th century (pp. 149- 155).

⁷⁰ There is suppressed time indexing going on in this example. These sorts of considerations raise interesting questions about how *particular instances* of relations of dependence are best characterized. But, I leave these questions to one side for now.

Thus, the formalizations and explication of grounding models all the intuitive features found in philosophical talk of metaphysical dependence. Hence, grounding is an accurate way to specify *what we mean* by metaphysical dependence. Crucially, what is being specified in this formalization is the notion of *dependence*, not how any particular entity manifests this metaphysical dependence relation. For example, to say that my pain is grounded in my neuro-receptors firing is not yet to say *how* it is grounded in them firing. Perhaps it is an example of strong emergence? Or perhaps functional realization? Further metaphysical work needs to be done to explain *how* things depend. I agree, to this extent, with grounding pluralists. Though I maintain along with the unitary theorists that there is but one notion of *grounding* which specifies the “dependence” side of metaphysical dependence, I agree with the pluralists insofar as I contend there are a plurality of ways to specify the “metaphysical” side of the *metaphysical dependence* relation.

Thus, grounding is a successful specification of metaphysical dependence that captures our use of that term more accurately than *modal* specifications in part because it builds into the relation certain features of *priority*—like anti-symmetry—that all back metaphysical explanations. This, however, does not yet recommend the grounding relation as the best account of dependence as such. For, there is another related notion that covers similar territory.⁷¹

Recent work in the logic of essence has clarified that, like metaphysical dependence, descriptions of essence cannot be merely modal. If so, however, then generic dependence might be definable *in terms of* essential dependence. That is to say, perhaps all cases of genuine dependence are *either* essential dependence relations or are reducible to essential dependence relations. This account has the virtue of explaining the formal features of grounding *in terms of*

⁷¹ It is also, often, argued that grounding is a *necessary* relation, especially if grounding characterizes *essential* dependence relations. This consideration will be discussed in more detail later.

the formal features of essence, which are arguably more intuitive and better understood. In the next section I consider this possibility by first describing essential dependence in general and then considering if grounding might be explained in terms of these essential dependence relations. I conclude that essential dependence is not capable of defining dependence in general, but that these notions are closely related. I thus conclude with my final taxonomy of grounding and dependence that I turn to apply back to my original definition in section 2.

3.3 Essential Dependence

In contrast to *existential* dependence, essential dependence involves identity or essence. Again, as Correia (2008) observes, this means that an essentially dependent entity is one that would not be the entity that it is (or be the way it is) had a condition of a certain type not been met (p. 1016). As with existential dependence the different conditions give different relations or properties of essential dependence. For example, Correia (2006) elsewhere distinguishes between the essential features that explain what it is *to be that type of thing* (what he calls the “objectual essence”) and the essential features that explain what it is *to be a certain way* (what he calls the “generic essence”) (pp. 753-754).⁷² In some ways, the recognition of a separate type of dependence is related to the critiques of mere modal covariance discussed above. Fine’s (1994) influential discussion about how modality is inadequate to model questions of essence utilizes the same examples. As such, following Correia (2008) I choose to formalize the existential dependence relation by turning to Fine’s (1995b) work regarding the logic of essence.

Fine (1995b) introduces an indexed sentential operator ‘ \Box_x ’ for ‘by virtue of the nature of $x \dots$ ’ or ‘ x is essentially such that \dots ’, so that given the sentence ‘ Y ’ we would read ‘ $\Box_x Y$ ’ as

⁷² To give a concrete example, to answer the question “what is Socrates?” is to answer a question about the “objectual essence”, whereas to answer the questions “what is it to be a wise person?” is to answer a question about the “generic essence.” Correia (2006) argues that the objectual/generic distinction *embodies*, as it were, the subject/predicate distinction in the realm of essence (p.754).

‘by virtue of the nature of x, Y’ or ‘x is essentially such that Y’. This allows us to simply and effectively state what is at stake in essential dependence. Once again we find complimentary specific and generic notions. First, what we might call *essential involvement*:

$$(5) \Box_x Rxy$$

[for some relation, R, x is essentially related by that relation to y]

$$(6) \Box_x \exists y (Fy \wedge Rxy)$$

[for some relation, R, there exists some ‘y’, y has property F and so x is essentially related by R to something that is an F]

Second, what we might call *essential necessitation*:

$$(7) \Box_x (Ex \rightarrow Ey)$$

[x is essentially such that it exists only if y exists]

$$(8) \Box_x (Ex \rightarrow \exists y Fy)$$

[x is such that it essentially exists only if there exists some y with property F]

These formalizations set the outlines of dependence relations as understood in the contemporary debates. The broadest disagreement, according to Correia (2008), is between “reductionist” views, which claim that essential dependence just is a modal requirement for existence, and “genuine essentialist” views, which claim that existential dependence does not entail essential dependence (pp. 1018-1019). These views differ with respect to which specific relations count as types of essential dependence. For example, reductionists will be comfortable thinking of supervenience relations as types of essential dependence, whereas genuine essentialists will argue that supervenience does not have the right kind of priority/explanatory asymmetry to be a real instance of dependence. As noted above, I think the arguments against a mere modal account of essence are decisive. Now, however, a new question reveals itself: what is the relationship between essence and grounding?⁷³

Many philosophers have, in developing Fine’s (1995a) early work on ontological

⁷³ For a recent overview of the historical development of these philosophical concepts and their relationship, including recent examples, see Ó Conaill and Tahko’s (2018) introduction to the special *Synthese* issue on ground, essence, and modality.

dependence, noted that there are deep similarities between the sort of essentialist priority relationships posited in Fine's analysis of essence and those present in grounding relations. This leads some philosophers (e.g. Zylstra 2018; Correia and Skiles 2017; Fine 2015; Correia 2013; Audi 2012, etc.) to suggest that grounding might be explained in terms of essence or that both are unified by some underlying commonality.⁷⁴ It would be beyond the scope of my dissertation to fully engage with these interesting and recent proposals; however, I briefly suggest some reasons why the two concepts should be kept distinct in light of Wilson's (*forthcoming*) criticism of Finean ontological dependence.

Wilson (*forthcoming*) argues that Fine's definition ontological dependence in terms of essence is lacking because it cannot make sense of priority relations in plausible cases. This is due to Fine's (1995b) particular logic of essence which has the following principle:

x depends upon *y* if *y* is a constituent of a proposition that is true in virtue of the identity of *x* or alternatively if *y* is a constituent of an essential property of *x*. (p. 275)

Put differently, according to Fine if *P* is true *in virtue of* the essence of an entity, *x*, then *x* depends on each constituent of the proposition expressed by *P*. The intuitive force of this principle comes from Fine's methodology of "real definition", by which the notion of one object depending on another is analogous to the nominal notion of one term being defined by another. Furthermore, it allows Fine to make sense of the distinction between existence and essence. By making the essence of *x* be identified with the collection of propositions (or properties) that are true in virtue of its identity, Fine explains how the *essence* of an object, rather than the object

⁷⁴ This, at least in Fine's case, is related to an underlying essentialist project that attempts to explain most of metaphysics in terms of essence. Fine (1994a) goes so far to suggest that necessity itself is perhaps explainable in terms of essence, noting that "metaphysical necessity [is] a special case of essence. For each class of objects, be they concepts or individuals or entities of some other kind, will give rise to its own domain of necessary truths, the truths which flow from the nature of the objects in question. The metaphysically necessary truths can then be identified with the propositions which are true in virtue of the nature of all objects whatever" (p. 9) For a recent argument against assimilating modal necessity to essence, see Wildman (2018).

itself, might depend upon another object. To give an example of this schema by utilizing the above principle:

‘{Socrates}’ depends on Socrates, if Socrates is a constituent of a proposition that is true in virtue of the identity of ‘{Socrates},’ or, alternatively, if Socrates is a constituent of an essential property of ‘{Socrates}.’

Does the definition give the correct reading in this case? It seems so. The proposition “contains Socrates as a member” is true by virtue of the identity of the singleton ‘{Socrates}.’ Further, Socrates is a constituent of that proposition. As such, we should say that ‘{Socrates}’ essentially depends on Socrates.

It is here, however, that Wilson notes a counter-example. Assume some sort of physicalism where lower-level physical entities, such as quarks, compose reality. Our best science suggests that quarks only exist in pairs or triplets, which compose non-fundamental entities like protons. Finally, suppose that it is *essential* to these quarks that they compose the very proton that they do. Now, quarks q_1, q_2, q_3 come into existence. So, in this example, the proposition “quarks q_1, q_2, q_3 compose proton P_1 ” is true by virtue of their identity and P_1 is a constituent of this proposition. Hence, by Fine’s principle, $q_1, q_2,$ and q_3 depend on the proton P_1 , but surely this is false. P_1 is a *less* fundamental entity than the quarks that constitute it, if anything P_1 should depend on them.

Of course, this shows at most that Fine’s analysis of dependence in terms of essence as a way of explicating essential priority is incomplete or requires modification. Wilson, however, diagnoses the problem in more detail. She claims that Fine’s central error lays in attempting to exclude entailments that were extraneous to the essence of the object by deploying the notion of something “pertaining” to the nature of the object and then analyzing “pertaining” in terms of dependence. As Wilson sees it, for y to not be extraneous to the nature of an object x —for y to

pertain to x —need not entail that x *depends* on y . Put bluntly, Fine seems to think that the essence of an entity, x , refers to the things upon which it depends, however, it is equally likely that the essence of x might refer to *what depends on x* . This suggests that a *full* account of essence is not able to capture dependence as such, since the essences of objects covers more types of relations than merely dependence.⁷⁵ Fine himself (2012a), recognizes that though these two concepts—metaphysical dependence and essence—are clearly related that is not sufficient to reduce one to the other. He makes the analogy to explanations of truth and explanations of identity, noting that though attempts to assimilate these two kinds of explanations are common, ultimately they must be taken separately. Fine (2012a) claims, “there is a similar error—but writ large over the whole metaphysical landscape—in attempting to assimilate or unify the concepts of essence and ground. The two concepts work together in holding up the edifice of metaphysics; and it is only by keeping them separate that we can properly appreciate what each is on its own and what they are capable of doing together” (p. 57).⁷⁶

3.4 Dependence, Grounding, and Causation

The above discussion introduced the notions of dependence and its grounding interpretation, as well as some of the controversies and potential problems with these

⁷⁵ Wilson herself seems to think that there might be a more general difficulty regarding how one utilizes schematic frames for these otherwise “thick” metaphysical notions. She recommends that we be pluralists and particularist about these notions, accepting only a minimal and nominal “family resemblance” between the concepts of “dependence” or “essence” that then is clarified on a case by case basis by more salient notions.

⁷⁶ Fine (2015), however, has recently reversed his position on this and taken on the challenge of unifying essence and grounding. He argues, roughly, that his earlier work did not appreciate the similarity between *explananda* in both essential explanations and grounding explanations. Once this similarity is appreciated, Fine (2015) argues, we can see the tasks of essence and grounding are not tasks with “distinct explanatory aims but are merely two different poles along a single explanatory endeavor” (p. 311). Put bluntly, Fine argues that generic explanation is *more* fundamental and thus serves to unify essence and ground. I do not have space to engage with this intriguing proposal, however, as a gesture towards a potential response: one might think that the plausibility of a kind of Finean union depends on whether you think grounding (and essence) are *metaphysically thick* notions, which involve entities other than facts, or not. Fine is more deflationary (in this limited sense) about grounding and thus can take generic explanation to serve a unifying role for the concepts themselves—rather than merely our methodological *use* of the concept. Those, like me, who are willing to bite the metaphysical bullet and accept a non-neutral *ontological* conception of ground (and essence), may justifiably demur.

conceptions. Furthermore, this detailed look at the logic and nature of grounding suggests an interesting taxonomy. Recall the progression of the dialectic: dependence is recognized as operative in many philosophical claims, attempts to characterize dependence in purely modal terms are found to be inadequate since they ignore *priority*, grounding is introduced to capture this prioritizing feature of dependence. Leaving aside the close relationship between essence and grounding, another similarity that came up repeatedly is the analogy between grounding and causation. Put bluntly, grounding captures the priority feature of dependence *in part* by modeling the features of causation, which *also* manifests a priority relation. This has been considered both an asset and a liability for grounding. On the one hand, it makes it similar to a widely accepted relation and thus more explicable to would be skeptics (e.g. Wilson 2018; Schaffer 2016a, 2016b, etc.). On the other hand, it raises suspicions that whatever grounding is it is too thin, too generic, or too derivative to be a primitive relation in its own right (Bernstein 2016; Koslicki 2016; Wilson 2014; etc.).

I believe the lesson to be learned is different. Grounding and causation share the same formal features because they are both species of *dependence*. In fact, dependence is the most generic notion.⁷⁷ In causation, current events *depend* on prior events, just as grounded entities *depend* on their grounds.⁷⁸ Hall (2004) recently argued that there are actually two concepts of causation: *causal production* and *causal dependence*. Leaving aside Hall's particular project, I think this does capture an important insight—causation differs from other forms of dependence by being *productive*. In general, the world is made out of dependences and, roughly, we can

⁷⁷ For an interesting alternative, yet related, account see Wilsch's (2016; 2015) account of grounding in terms of nomological determination—that is to say, determination by “laws of metaphysics” analogous to the laws of nature.

⁷⁸ I am here uncritically assuming that causation always involves a temporal priority relation such that the cause (temporally) precedes its effect. There are reasons to doubt this, moreover, I take it that depending on how one answers this question one arrives at a different conclusion regarding the relationship between causation and dependence as such. I do not, sadly, have space to address the complexities of these arguments here.

speak of these as metaphorically running along two axis: diachronic relations that explain how the world “unfolds” and synchronic relations that explain how the world is “built”. Given this taxonomy, grounding’s unique logical structure and internal relations are primarily capturing the generic features of *dependence* that it shares with causation, this is why the logic of both are so similar.⁷⁹

What then makes grounding unique? This taxonomy suggests at least *one* unique feature is that it is *non-productive* in the relevant sense. It does not essentially involve the kind of concrete dynamical processes that are widely held to inform causation (i.e. transfer marks, mass-energy, or some other conserved physical quantity, etc.). It is, instead, metaphysical and thus non-causal. This feature of grounding will become important later in discussing its relevance to action. For, it is initially puzzling how something that is defined by being *non-productive* might explain or define something that is paradigmatically a “doing” of agents.

Once this broader taxonomy is recognized it becomes much easier to see how the *specific* relations of dependence that are non-productive might all be characterized as a species of ground. Other candidate relations that are thought either to be involved with or subsumed under grounding relations include: constitutive relations, compositional relations, determinate/determinable relations, part/whole relations, set membership relations, and proper subset relations, to name a few. There is little consensus regarding how these various relations should be conceived of taxonomically. As such, I set these debates aside, instead roughly characterizing non-productive dependence using the grounding relation.

⁷⁹ Alternatively, Bennett (2017) provides a comprehensive argument for taking the concept of “building” seriously as a unified *family* of relations.

4. Action as Essential Metaphysical Dependence-Redux

With the detailed considerations of dependence, essence, and grounding in hand, I return to my initial rough definition:

Action as Essential Metaphysical Dependence (ver. 1): A mental event e is an action iff e essentially metaphysically depends on interactive mental properties, P .

With the concepts used in the previous section we can now specify the generic essential metaphysical dependence relation more precisely:

Generic Essential Metaphysical Dependence (ver. 1*): For entities x, y ; y essentially metaphysically depends on x , dff: $\Box_x Rxy$, where $R = Xs // Ys$, $[x \in \{Xs\}] \wedge [y \in \{Ys\}]$, and $y = \Box_y \Gamma$, hence $\Box_x(x // \Box_y \Gamma)$.

Or more directly:

Generic Essential Metaphysical Dependence (ver. 2*): For entities x, y ; y essentially metaphysically depends on x , dff: $\Box_x(x // \Box_y \Gamma)$.

I am here utilizing both Fine's essential operator ' \Box_x ' which means "by virtue of the nature of x " and Schaffer's binary full grounding predicate ' $Xs // y$ ', which means " Xs fully ground y ". So, informally, to say y essentially metaphysically depends on x means "by virtue of the nature of x x grounds those properties brought about (or propositions made true) by virtue of the nature of y ". Furthermore, since this is meant as a generic formalization of the relation I take onboard Dasgupta's (2014) claim that *full* grounding should be, in the first instance, formalized as a many-to-many relation. Of course, this does not preclude a full grounding relation between just two entities, since the limit case for a set of Xs is *one* x . It does, however, mean that many-to-many full grounding relations are not ruled out by formal fiat.

This formalization highlights a crucial ambiguity in my original informal statement. In that statement, "essential" was ambiguous between the essence of x and the essence of y , this formalization shows that I mean *both* though in different ways. I claim that x is essentially

related to *y*, which Correia (2008) calls “essential involvement.” I stipulate this essential relation is best understood as *grounding*. Following Finean style semantics, I claim that *y*’s essence, understood as the list, ‘ Γ ’, of properties brought about (or propositions made true) by virtue of *y*’s nature, is one of the *relata* of the grounding relation. So, “essential metaphysical dependence” is “essential” in two ways: (i.) it is part of ‘*x*’s *essence* that it ground *y*, and (ii.) one of the *relata* of the grounding relation is the *essence* of *y*.

We can now apply the specification of essential metaphysical dependence to the specific case of action:

Action as Essential Metaphysical Dependence (ver. 2): A mental event *e* is an action iff by virtue of the essence of interactive mental properties, *P*; *P* fully grounds all the properties brought about (or propositions made true) by virtue of the essence of *e*.

Two more minor clarifications. As Fine (1994b) notes we can distinguish between the constitutive essence of an entity and its consequential essence. Roughly, the constitutive parts of an entity’s essence are those properties that are not had by virtue of any more basic essential properties of the entity; otherwise it is merely a consequential part of the essence. This is important to prevent an infinite chain of entailments becoming part of what defines an entity. To give an example, it is constitutively essential to my volition to drink coffee that it be intentional, however, it is only consequentially essential to my volition that, the disjunction, ‘it is intentional or not’ is true. Both claims are made true by virtue of the essence of volitions, however, only the first forms the “core” of what it means to be a volition (what Fine calls “real definition”), whereas the second is a logical consequent. Second, if we specify the essences of things by virtue of what properties they possess (or what propositions they make true) then we still need a way to limit the range of the consequentialist essences. For example, if we do not limit essential consequences at all then *any* necessary logical consequent would be essential as well. But, it

seems incorrect to say, for example, that $2=2$ is part of even the *consequentialist essence* of my volition to drink coffee. This is where Fine (1995a; 1995b; 1994a) introduces the notion of “pertaining” to restrict notion of essence. Though many statements might be true with respect to the essence of x as a matter of consequence, only a subset of these will actually *pertain* to the essence of x , even understood consequentially. As noted above, Fine’s particular account of pertaining appeals to considerations of “dependence”, which leaves him open to counterexamples suggested by Wilson, among others.

I cannot motivate a full account of what it means to pertain here in such a way to appropriately limit the consequential essence of an object. Nor will I adopt the Finian “neutral” methodology of attempting to describe essences wholly in terms of their consequentialist results and then extrapolating the constitutive essence via “abstracting away” the merely consequential bits. Though I agree with Wilson’s assessment that, in many cases, defining metaphysical terms in the most theoretically neutral manner is helpful for making dialectical progress, this project is already specific enough to warrant defining particular notions. All I need for the theoretical outline is the notion of constitutive essence and an informed yet intuitive sense of “pertaining.”

This, then, gives the final version of my account:

Action as Essential Metaphysical Dependence (ver. 3): A mental event e is an action iff by virtue of the constitutive essence of interactive mental properties, P ; P fully grounds all the properties brought about (or propositions made true) by virtue of the constitutive essence of e .

With this account fully specified I now turn to examine its applicability and success. I do this in two ways. First, I examine how it addresses the considerations brought up in chapters 1 and 2. Specifically, can it meet the desiderata of standard non-causalists while addressing causalist concerns. Second, I turn to address the specific problems that my account generates regarding circularity and control. Finally, I conclude by pointing out a final metaphysical concept—

emergence—that is necessary for my account of action to avoid both problems of transitivity and also fit in a causally structured world like ours, I then take up this concern in chapter 4.

5. Action as Essential Metaphysical Dependence—Application

Given my motivation for crafting a view that (1) combines the insights of both agent-causal and non-causal views and (2) responds to causalist objections to non-causal views. How does essential metaphysical dependence fair in comparison to other theories? First, I examine how account synthesizes the agential and non-causal views by satisfying the desideratum of non-causal views while still positing a direct extrinsic connection to the agent.

5.1 Satisfying Non-causal Desideratum

Essential metaphysical dependence satisfies the most central commitments of non-causalism. First, according to Ginet, as well as Goetz and McCann, an essential feature of our acts is their *active* quality. This is one of the reoccurring critiques non-causalist raise against causal theories. The strongest version of this view is held by Goetz. By his lights, to be caused conceptually entails being passively produced. Actions are intrinsically active; hence, they cannot be caused. At first glance, this seems to be a difficulty for my account. Dependence relations sound like passive relations, in part because the concept of “activity” suggests “production” and generic dependence relations are neither causal nor agential and thus may intuitively seem not “productive” in the relevant sense. This concern, however, is unfounded. To show why I first must clarify what is conceptually at stake in the active vs passive distinction.

What does it mean to say that something is active as opposed to passive? Roughly, something is *active* if it “does” things, something is *passive* if it has things “done to” it. These are already metaphysically thick notions, but some obvious conceptual points should be highlighted. Explanation, as such, is more abstract and general than the active/passive distinction. I might

explain something by citing that X is active with respect to Y, but not all explanations need to fall into this category. With this in mind, recall that the non-causalist critique, especially for strong non-causalist views like Goetz, involve pointing out that if simple mental acts are *intrinsically essentially* active then it would be odd to say they are produced. Why? Because to be produced is to be “done to” and such mental acts are essential things that “do”. Given the plausible claim that an entity cannot have a conflicting essence it seems that such mental entities *cannot* be the sorts of things that are “done to”, which is to say they cannot be passive.

Earlier, (in chapter 2), I criticized Goetz’s view on the grounds that he confuses the constitutive conditions for something’s essence with the conditions for something’s existence. This critique still applies, but I wish to highlight a difference concern in order to explain why the active/passive distinction is not a difficulty for essential metaphysical dependence.

Essential metaphysical dependence *explains* the active nature of simple mental events but this explanation is not *productive* and therefore does not entail that the event is “done to” and hence passive. Explanation is a more general category than production and need not entail that the *explanans* is passive with respect to *explanandum*. Hence, to say that mental event *e* is an action iff *e* essentially metaphysically depends on interactive mental properties *P*, does not conceptually undermine the intrinsically *and* essentially active character of *e*. In fact, it aims to *explain* that character in terms of its being grounded in interactive mental properties.

A generalized example of this claim elucidates this point. Water molecules are essentially grounded in a certain structure. Further, water’s property of being a solvent is *because* of its essential structure. It would be odd, however, to say the *event* of salt dissolving in water is passive with respect to the structure. Such an event is an active manifestation of one of the essential properties of water—loosely speaking, the water molecule *does*, the salt crystal is *done*

to. This active manifestation, however, simultaneously essentially depends on the structure of water molecules. In like manner, even if simple mental events essentially depend on or are grounded in interactive mental properties this does not conceptually entail that simple mental events are passive. They can both depend on these mental properties and be intrinsically active, indeed my claim is that their intrinsically essentially active character is *explained* by their dependence relation to these mental properties.

This raises a second concern. Ginet, among others, argues that whatever makes a simple mental event an action must also be *intrinsic* to the simple mental event. In arguing against agent-causation Ginet points out that if agent-causation were true mental actions should have a complex structure. The act, as a whole, would consist of two parts: the agent-cause and its effect, the simple mental event. This dipartite structure would be what the act consisted in. But it seems, he argues, that we recognize the difference between a simple mental act and a mere mental event without reference to the *relation* that obtains between the agent and the act. This criticism has some weight against my view. If a simple mental event is an action by virtue of being essentially dependent on mental properties of the agent there seems to be a similar complex structure between these properties as ground and the simple mental event as the grounded. If so, then the same critique applies to essential dependence.

This critique, though challenging for my proposal, can be addressed. Ginet's argument, like McCann's and Goetz's, is in the first instance epistemic. We differentiate between simple mental acts and other mental occurrences because of some intrinsic features rather than by introspecting an extrinsic relation between the mental event and the agent or a property of the agent. As Ginet (1990) notes, a causalist might claim that the intrinsic phenomenal features are just an intrinsic phenomenal sign of the extrinsic causal relation (pp. 9-12). He dismisses this

kind of response for leaving difference between veridical and illusory instances of the phenomena unexplained. If the agent-causalist responds that the phenomenal sign is *always* veridical this seems *ad hoc*. Furthermore, such a position, Ginet (1990) argues, would be needlessly complex. If we only attend to the intrinsic phenomenal features to distinguish between the two events why posit a further relation?

Regarding the first concern, on my account saying the intrinsic quality is a veridical indication of the essential dependence relation is not *ad hoc*. This is because there are no deviant dependence relations in the same way there can be deviant causal chains. In particular, since it is *essentially* dependence, plausibly, the conditions for something's essential ground must exist if the grounded thing exists. I agree with non-causalists that part of the essence of simple mental acts is a certain kind of intrinsic *active* quality, but I add that this quality, insofar as it is essential to the mental act, is itself *grounded* in the interactive mental properties of the agent. If the actish phenomenal quality is present the essential grounding relation *must* be present; if it were not the simple mental act would not be what it is essentially. This response is not available to agent-causalists because it is possible, according to them, that actions are sometimes caused by something other than the agent and still feel like actions. On my proposal, it is open for the grounding relation to always explain the essential intrinsic *actish* qualities of the simple mental event even if there is another causal explanation for its existence.

Regarding the second concern, the purpose of positing a further relation is, in part, to address the causalist criticism that non-causalism's appeal to phenomenal or experiential properties are naturalistically inadequate. As O'Connor (2000) says, "seemings are not sufficient for the realities" (p. 26). The standard non-causalist response is to (1) point out that non-causal views are consistent and are not proven wrong by merely disputing their plausibility and (2)

point out the difficulties with causalist alternatives. While this is an adequate defense, it is clearly not the most persuasive. Indeed, it is this kind of argumentation that leads to the dialectical stalemate in which non-causalist and causalist theories are enmeshed. If one could propose an alternative conception that addresses this causalist concerns about experiential evidence, while at the same time meeting the non-causalist desideratum this alternative would be dialectically superior. My account is intended to be just such a view.

These considerations show that my view meets the central requirements of a non-causalist view. Furthermore, it incorporates the central causalist insight regarding the need to ontologically ground actions in some feature of the agent, in my account the interactive mental properties that jointly constitute the agent's reason-responsive mechanism. How does this account address the central concerns causalists raise against non-causalist views? Again, I focus first on attempts to answer the question, what *makes* a mental event an *action*, not questions regarding what *explains* action. I address that second question below. However, this distinction should not be overdrawn. Some causalist views collapse these answers together such that what makes an event into an action *also* provides the action explanation. That being said, making this distinction allows us to focus on the most salient features of the view.

5.2 Addressing Causalist Concerns

A common objection, then, to the non-causal definition of action is that such a definition does not fit with the causally structured order of the natural world. Seemings are insufficient for realities and since all of the non-causal views either rely on such seeming to be what makes an event *into* an action or are evidence for what does make an event into an action, there is a fatal obscurity at the heart of such non-causal views. This initial worry can be bolstered by parsimony

arguments. For example, the natural world is causally structured thus it seems *ad hoc* to suggest that humans are somehow exempt.

These concerns are, so far, intuitive. However, Mele (2017; 1992) and Clarke (2010) have tried to make them more precise. I have already discussed in some detail these objections to non-causalism in chapter 2. Though the examples given are primarily intended to undermine the *action explanation* side of the discussion rather than the *action definition* side, the point still goes some of the way toward making the above concern clear. Consider the following case:

Sarah intends to raise her arm in order to get the attention of a speaker. Unbeknownst to her, the neuroscientist, Andrea, has crafted a device that allows her to remotely manipulate Sarah's brain. With this device she both makes Sarah bodily raise her arm and *actively will* to raise her arm. Andrea is doing this with the intention to test her device. Throughout this event Sarah maintains her intention to raise her arm in order to get the attention of a speaker.

What is the nature of the action in this case? According to causalists like Clarke and Mele, even though Sarah's overt bodily movement was initiated by a willing with an actish phenomenal quality and has an accompanying intention, it would not count as *her* action, rather it is Andrea's. It is explained by Andrea's intention, not Sarah's. These kinds of counter-examples, which were discussed in the previous chapter, highlight the non-causalist difficulty explaining what the *salient* connection is between the non-causal features that are meant to explain action and the *actions themselves*.

In the previous chapter, I claimed that these cases offer a direct counter-example to views like Ginet's, which allow for our actions to be caused-events if they are not actions *by virtue* of being caused. It does not; however, undermine other non-causal views. This is because theorists, like Goetz for example, fall back on "begging the question" style responses. These proponents of "strong non-causalism" believe that it is conceptually impossible for our active willing's to be caused, hence the counterexample does not succeed. This response, though conceptually

available, is unsatisfying. A causalist would, I believe rightly, see this as a last resort of the non-causalist to save their theory regardless of the weight of evidence.

How does my view fair against such critiques? In response to the intuitive critique my account fairs rather well. Seemings are not *sufficient* for realities, but if one can establish that a seeming is essentially grounded in a property of mind, then this phenomenon becomes veridical indicator of a real extrinsic connection. The relation of essential metaphysical dependence forges a connection between a mental property of the agent, in fact the kind of properties that we might rightly call *intrinsically agential*, and our simple mental events of willing, deciding, and forming intentions. In turn, these events get their unique phenomenal character from their essential relationship to that on which they depend, in the same way that the macroscopic properties of water are grounded in the more fundamental structural properties of the atoms that constitute a water molecule.

The more formalized version of this critique raises another issue, namely, just what is the relationship between the various parts of the mental ontology such that they define what actions *are* without becoming a causalist view? Or, to put it another way: Can my version of non-causalism do justice to the intuition that Sarah does not act when she actively raises her arm under the control of Andrea, rather, Andrea acts *through* Sarah? These questions push us towards addressing not just what makes an event into an action but also what *explains* actions. Which is to say, what role do reasons play in action theory.

There are really two questions here. *How* do reasons (or intentions) explain actions and *who's* reasons (or intentions) do they belong to. This difference has not been attended to sufficiently. Causalists, like Clarke, propose that the best explanation for our intuitions in the neuroscience cases is that the intentions of the neuroscientist are enacted by the victim's action,

they causally explain what is, in fact, going on and for that reasons we should doubt any non-causalist proposals. What has been lost in this discussion is the question of to whom the intentions *belong*. To see why this is important it is crucial to become a little clearer on the relationship between mental events like volitions, decisions, intentions, etc.

First, to recall some basics, all conscious mental events (and perhaps many unconscious mental events) are made up of (at least) two features: content and mode. The content of a mental event is what the event is, to speak loosely, “about”. In contrast, the mode is *how* the content is held in the mind—for instance, as a belief, or as a desire. This, following McCann, is helpfully schematized as: Mode (*Content*). Consider my having an occurrent conscious belief that it is about to rain. This mental event I am experiencing is describable as: Belief (*that it is about to rain*). Or, in like manner, suppose I form an intention by deciding to go for a run later. This mental event is describable as Decide (*to go on a run later*) and this, in turn, leaves me with the mental state of Intend (*to go on a run later*).

Again, I emphasize that this is only a schema or model of our folk psychology. Within this framework, however, we can more precisely indicate many useful lines of inquiry. For my purposes, I merely highlight two features. First, that the *mode* by which a mental event is held in a mind is the most important for understanding function or psychological use of these events. Second, these modes are *types*. Their tokens are the particular instances of mental events that are united with particular content.

What then, more particularly, are we to make of intentions? McCann calls them “states” because they are temporally extended and thus different from mental events, like deciding. This is a little idiosyncratic, many philosophers regard states as merely a *type* of event that has certain

features (being temporally extended, having no clear culmination, etc.).⁸⁰ Regardless of the best way to parse up the terminology, intentions are distinguished by two features. On the one hand, they are distinguished from willings or decisions by being temporally extended *plans* that may or may not be enacted or realized at a given time. On the other hand, they are distinguished from mere desires or desire-belief pairs by being *practical*. That is to say, to intend something is more than merely having a belief about how to do something or even a desire to do the thing, it is to have a commitment to the action. Paradigmatically, this means that when one intends something that means (1) there was a process, however brief, of deliberation which formed one's desires and beliefs into an intention, and (2) the intention will, normally, be realized in some sort of action.

Also, volitions, mental events of willing, are intrinsically intentional. That is to say, when I will to do something on a whim with no prior thought or deliberation this willing is still an intentional act of mine, even if there was no prior intention. I cannot, as it were, accidentally will something. Now, of course, the results and consequences of my action might be unintentional but that is because of a failure regarding the action and its result, *not* a failure internal to the willing itself. If so, then there is a question. What is the relationship between a willing and an intention. McCann suggests, I think rightly, that the relationship between them is one of ratification. The

⁸⁰ Indeed, there is a larger question here about what exactly a generic "state" is as opposed to an "event". There is no *clear* consensus, however, it seems that for most philosophers the terminology of "events" and "states" are used to indicate the way objects change or stay the same. Roughly, events are changes in objects or involving objects, whereas states indicate an object or something involving an object having a property (or properties) without change. An object can, of course, change from one state to another state, but that change itself is an event. For example, I might be in a state of unconsciousness while sleeping, then the event of my alarm going off wakes me up, now I'm in a state of consciousness. As this example shows, events and states can both co-locate and constitute each other in complex ways. So, for example, my state of unconsciousness while I'm sleeping is partly constituted by repeated instances of the event of my lungs filling with air and expelling air. Conversely, my being in the state of living partly constitutes the state of the earth having human life, though if an asteroid is about to hit earth this large-scale state will change because of that, so-called, extinction level event. Given this, the states and events that are salient for a given explanatory question depends, partly, on the scope of inquiry.

willing takes the content of the intention and “affirms” it *by so willing*. This conceptualization is because, in part, we must avoid the absurdity of there being a second set of intentional content. Intentions are never doubled. Of course, one might have multiple *goals* that one aims at with one and the same action. However, it borders on incomprehensible to speak of intending to pick up a book and willing, with its intrinsic intention, to pick up a book and then having both these intentions exist side by side. It is as though one could form two beliefs that it was about to rain, so that just in case the first belief failed you would have a back-up.

So then, when I *enact* or *realize* an intention what is happening *just is* that by willing I take up *as the intentional content of the willing* the prior intention I had formed, which then guides my volition. This, then gives us a response to Clarke. For we can ask, what was Andrea the neuroscientist causing? If she was *causing* Sarah to will to raise her hand then this must be an intrinsically intentional action, if so what was its content? If it was a ratification of Sarah’s prior intentional state then there is no puzzle. Sarah did (unfreely) act on her intention to bid for the painting. If there is any puzzle here it lies in the free or unfree status of her action, indeed a case like this is more analogous to Pereboom’s manipulation style arguments against compatibilism. If, however, Andrea causes Sarah’s volition to have a different intrinsically intentional content that runs contrary to her prior intentional state there is still no puzzle. For, it is sufficient to say that the intentional content did not belong to Sarah, instead it is Andrea’s intentions that explaining, fully, the relevant action and Sarah is merely a tool for this action. To put it more bluntly, this sort of picture is one where Sarah would have seemed, from her own perspective, to have chosen accidentally, which is to say it would be a case where *Sarah* wouldn’t have willed anything at all.

But, it cannot be the case that Sarah wills something and simultaneously has an intention that both accompanies and is disconnected from the willing. To will *just is* to act intentionally and the intentional content is *part of* the willing. It is not doubled by the willing. This conception of the relationship between willing and intentional content allows my view to forge a connection between the reasons and the willing itself, thus avoiding the causalist objection that non-causal conditions are too thin to define or explain action.

This response also highlights my account of reason explanation. To act *for* a reason just is to have that reason figure in the constitutive essence of the relevant mental event, wherein it functions to teleologically explain *why* one acted as one did. When I choose to raise my hand to order to ask a question, the reason “to ask a question” is the goal or purpose for which I raised my hand. Moreover, it is essentially part of *that* willing since such volitions are intrinsically intentional. I may indeed have many reasons for doing what I do but the one that explains my so acting is the one that (a) teleologically explains my action and (b) essentially figures in the relevant mental event.

How does this fit with my overall essential metaphysical dependence account? Recall that the interactive mental properties are, plausibly, constitutive of the reason-responsive mechanism of minds. If so, then these properties are *essentially* the sorts of things that ground the essential nature of the basic mental acts, like volitions. Speaking loosely, what it means for these properties to be reasons-responsive *just is* that they ground those mental events that (a) have reasons as a constitutive part of their essence, (b) in a functionally teleological manner. This conception of reasons relationship to simple mental acts answers the two central causalist questions. It answers the question regarding how reasons *explain* action in a manner common to most non-causalists—teleologically. It also, however, answers the question how are reasons

linked to the actions they explain. This second question is at the heart of the causalist objections raised by Clarke and Mele. My view answers this question by noting that the reasons for which we act are not a sperate entity, such that they might be explanatorily disconnected in intervener cases. Instead, they are *constitutive* of the volitions that they explain and they are made constitutive by being *grounded* in those interactive properties that constitute our reasons-responsive mechanism. Hence, both the constitutive relation which explaining reasons have to the actions they explain, and the grounding relation, which explains how these reason laden mental acts come to have they character they do, are non-causal. Hence, reasons can non-causally explain actions without falling victim to intervener style cases.⁸¹

6. Objections to Essential Metaphysical Dependence

The previous section presented the final version of my account and then examined how it met both non-causalist desideratum and causalist objections. I now examine two difficulties for my argument—the *circularity problem* and the *control problem*. These objections concern the nature of the grounding relation and the nature of mental properties. As such, though they

⁸¹ At first blush, it might seem that my view excludes the possibility of acting for no reason at all or “on a whim.” I do think that it is possible to act without *any* prior intention or rational plan. Moreover, I hold that it is possible for an agent to act *suddenly* on a whim. I do not believe, however, that such actions are without reasons—broadly construed. Such actions are not *non-rational* like a reflex nor *irrational* like a compulsion. Instead, I agree with McCann that to will anything requires that there be content to the willing—something that the willing is directed towards. As I noted in chapter 1, this is what McCann means when he says that willing is *always* intentional even if it lacks an accompanying intention. We never will accidentally. The upshot of this is that to act “on a whim” is not the same as acting *accidentally* or *inadvertently*. Moreover, this content is what makes such actions rationally explainable, in a broad sense the intentional content is our reason for acting. In fact, if you press someone who genuinely acted “on a whim,” they often merely cite *the thing they did* as the reason for doing it (e.g. “I just did”, “just because it was there”, “it seemed like a good idea at the time”, etc.). I think this kind of response is an instance of sincere reason giving but the reasons given are the *barest* possible—a mere restatement of the content of the willing itself. Of course, for most *practical* purposes we talk of these sorts of reasons as “reason-less” in some sense, but we should not let this practical function obscure the metaphysical point. This mirrors Goertz’ claim that so-called “irrational” compulsions are better understood as instances where the weight of reasons is subjectively thrown *wholly* on one side of a given choice. Such compulsions are properly called irrational, Goertz thinks, because we understand that the agent is not appreciating their reasons correctly not because the agent lacks reasons. In most cases, the compelled agent is still acting *for* reasons, however, their reasons are weighted unusually. My view, then, makes this same point in the other direction. If compulsion is a case where reasons are *excessively* on the side of one choice (thereby practically irrational), then “acting on a whim” is a case where the reasons are *deficient* enough to reduce to merely the content of the willing (thereby practically reason-less).

present challenges for my account, they also allow me to clarify what makes essential metaphysical dependence distinct from other accounts of action.

6.1 Circularity Problem

My account might appear to be subject to a vicious circle. I define interactive mental properties functionally, as being those properties that give rise to our ability to interact with the world. Moreover, I claimed that these properties are constitutive of our agency. Even more directly, these properties were defined by how they ground events with the following two features: (1) they involve interactions with the world (e.g. volitions), and (2) they are partly constituted by the reasons for which they are done. I also argued that simple mental events are actions just in case they are wholly grounded in these interactive mental properties. If so, then it appears I am saying interactive mental properties are essentially defined by grounding simple mental actions *and* simple mental actions are essentially defined by being grounded in interactive mental properties. The circle is complete and it is unfortunately quite small. This objection fails, however, by equivocating on the different ways we can essentially define entities. I first clarify the objection by contrasting it with a similar objection raised against dispositional monism. Second, I show how the objection fails by confusing the scope of essential definition.

Initially, this objection appears like a concern regarding circular definitions.⁸² When defining something the *definiens* gives meaning to the *definiendum*. When definition works it is because the *definiens* is independently meaningful so it can “pass on” its meaningfulness and thus elucidate the target term. This motivates the thought that definition must not be circular,

⁸² Much of my discussion here is indebted to Bird’s (2007) work on regress objections to monistic dispositional essentialism. Where I am quoting Bird directly I will indicate, but in general this whole section should be understood as inspired by Bird’s arguments.

since at some point there must be a term that is already meaningful or acquires its meaning directly (e.g. by ostentation).

The criticism of my account applies this same structure but to “real definition,” which defines not the meaning of a term but the essential identity of an entity. I am, after all, answering the question of what makes something the kind of thing that it is. Given this project, the worry is not that *meaning* is constantly deferred but, that *essence* is being deferred. The ontological buck passes back and forth between the mental property and the event it grounds without either gaining enough *reality* to be adequately defined.

This worry shares many similarities with critiques of monistic dispositional essentialism—the view that all fundamental properties are essentially dispositional. A property is essentially dispositional if it is defined by its causal profile and the conditions under which this profile manifests, sometimes called the *manifestation* and *stimulus* respectively. So, for example, if the property of fragility were essentially dispositional this would mean that fragility *just is* understood as the conditional “if struck (stimulus) then shatters (manifestation)”.⁸³ Given this metaphysical framework dispositional monism faces the same ontological circularity concern as my account of action. The reality of any given property, for the dispositional monist, is given by citing the two properties involved in its stimulus and manifestation. Each of these properties, however, are themselves dispositional on a monistic view, therefore they are defined in terms of another set of properties, and so on. So, either dispositional properties are infinite or at some point they make reference to properties that have previously occurred and are thus circular.

⁸³ Of course, “fragility” is merely a toy example. It is almost certainly not a fundamental property and thus does not fully fit the monistic dispositional essentialist analysis, however, it is adequate for explaining the general idea of a dispositional property.

To be clear, if the concern is merely about ontological *reality* it may be that this is unfounded, for it clearly begs the question. Why think that dispositions have less “reality” than other accounts of properties? As Bird (2007) argues, the primary categorical account is, if anything, thinner than the dispositional account (pp. 521-522).⁸⁴ It is plausible to think, however, that the criticism concerns not reality exactly but *identity* or *essence*. How can dispositional properties have the *essence* or *identity* that it does if that is only given in terms of other properties that then also have their essence or identity in terms of other properties, and so on. No single property has its identity fixed since it gets its identity in terms of another property (Bird, 2007, p.523). This version of the criticism is clearly analogous to the one raised against my view.

Bird’s solution involves appealing to how the identity of individual properties can be non-vacuously given by reference to the overall interconnected structure of dispositions via well-founded graph-theoretic principles. If the structure of dispositional properties follows these constraints it is, in principle, possible to define the identity of every property in the structure even though they are all interconnected. This attempt is intriguing, and I think successful, but it is not helpful in my case since I am not implicating the entire structure of property relations. This response, however, does highlight that my critic misunderstood the ontological scope of my account. It is the monism of Bird’s account that generates the problems of identity. In contrast, the definitions involved in my account of action do not rest on an underlying monistic conception, nor are they even about the same *kind* of entity. To show why, I consider each side of the “vicious” circle in turn.

⁸⁴ For, on the categorical account, properties only possess the features of being: (1) distinct from other properties, (2) universal and thus have instances, (3) for some n it is an n -adic universal. But, all these features are possessed by dispositional properties as well in *addition* to the claim that they have an essentially dispositional character.

First, regarding the interactive mental properties, as noted in the earlier discussion of essence it is reasonable to think that the essences of things might rest not just on what they depend on but also what depends on them. This is clearest in Wilson's example of fundamental particles. In her case, the particles' essence, understood in terms of those propositions made true by virtue of its constitutive identity, only makes references to those things that it grounds. I add to this account that we might speak of fundamentality in a restricted sense as regarding that which is fundamental to a domain of objects. For example, suppose physical reductionism regarding mathematical objects is true. Now, suppose that all mathematical structures, and so the entities and operations that are defined by them, could be derived from some small set of axiomatic functions (like the successor function), ' F '. It seems that the following two claims would be true: (1) F is mathematically fundamental and so is defined in terms of what it does or grounds, (2) F is dependent on physical reality, assuming the truth of mathematical reductionism. So then, these two examples mean that, first, it is intelligible to speak of fundamental entities as being essentially what they are in reference to what they ground, second, it is intelligible to speak of domain specific fundamentality.⁸⁵

Certain mental properties, like the interactive mental properties that constitute our agency, are fundamental with respect to the domain of agential minds. If so, then it would be understandable for them to be essentially identified with those mental events that they ground since, insofar as we are defining their essence *qua* mentality, there is no *more fundamental* mental entities to define them in terms of. If this picture is correct, then there is nothing unintelligible about a fundamental mental property being essentially defined by the mental

⁸⁵ I have presented this conception of domain specific fundamentality mostly by example. A more rigorous definition would, I think, define it in terms of domain specific explanatory failure (e.g. the mathematically fundamental objects cannot be explained any further *que* mathematical explanation). I, however, do not have the space to explore what a full account of domain specific fundamentality would look like here.

events associated with it.⁸⁶ But, even if it is intelligible to define a fundamental mental property in terms of the events it upholds, the problem lies in how the events are *also* defined in terms of the property that grounds them. It was a vicious *circle* after all.

This objection also fails. The grounded mental event is not ontologically akin to the property that grounds it, both by virtue of being a different kind of thing—an event not a property—and by having a different kind of status—derivative not fundamental. This means that what is crucial for the essential definition of the mental event in question is how it is an *action*. The definition is only insofar as it counts as an action is it grounded essentially in interactive mental properties.

Another example might help. Suppose that Descartes was right about people being fundamentally “thinking things”. Plausibly, what people are would be essentially defined in terms of being the sort of thing that thinks. How are thoughts defined? It seems that, almost trivially, thoughts are going to be essentially defined as the sort of entity that are brought about by minds (i.e. thinking things), however, this offers no special puzzle of definition because “thoughts” have many other properties that aid in their definition. That is to say, the essential definition of thoughts is not exhausted by referencing their dependence on thinking things *even if* that feature of their essences is in some ways most determinate or constitutive.

In like manner, the simple mental events’ essence, which is essentially grounded in interactive properties of mind, is defined both by being grounded in this way and by other constitutive features of its essence. For example, volitions may essentially have an actish phenomenal quality (as Ginet suggests) at least when instantiated in self-conscious minds like

⁸⁶ Indeed, I already ran a similar gambit earlier when discussing the property of consciousness. Insofar as we are willing to have consciousness defined in terms of the phenomenal awareness (of various types) of events that depend on it.

ours. The *actional* nature of volitions is defined by virtue of its full grounding in interactive properties and this is indeed essential. A volition doesn't *accidentally* become an action, it *is* an action by virtue of being grounded in the appropriate fundamental mental properties; however, to say that the essence of a thing is fully grounded and that this specifies the *type* of thing it is does not mean that the identity of the thing has been exhaustively given. Just as in my example, where the *type* of thing, "thoughts," were wholly grounded in "thinking things," but were further specified by other features. *Of course*, in a sense, being in the category "thought" was perhaps the most encompassing or constitutive part of their essence just as counting as an action might be the most constitutive part of the volitions' essence *but* constitutive is not the same as exhaustive and thus we can avoid the essential circularity objection.

At this point someone might, rightfully, object that even if this response addresses the concern regarding circularity of "identity" I have not address the similar concern of conceptual "emptiness." While it is perfectly *intelligible* to understand interactive mental properties in terms of their functional role, or what they ground, it is also unsubstantial. Insofar as the only difference between interactive and responsive mental-properties is given in terms of the kind of mental events they ground I have not really said what they are or why I am justified in positing their existence.

To answer this understandable concern I refer back to section 2 where I originally introduced the interactive mental property. Part of my argument's plausibility lay in its analogy to the property of consciousness. Insofar as the "interactive mental properties" are meant to be situated to agency just as the "responsive mental properties" are situated toward consciousness, I do not think the objection that my argument is empty holds. Both our philosophical and folk psychology concepts *already accept* a property like this as the unifying and explanatory ground

for agential mental events.⁸⁷ My account is only novel in detailing the relationship *between* these events and their ground. I claim this *relation* is what makes the difference between an action and a non-action and this relation is non-causal. Furthermore, as shown above, my account of this relation is substantive and highlights the *priority* of the agential side of the relation—hence, it is an agency-first account. Given this, it seems that the more pressing objection for my account is not about circularity, but rather about how the relation *itself* presents problems for giving an adequate explanation of action.

6.2 Control Problem

This leads to the second problem—*control*. Our simple mental acts are paradigmatically the sorts of things that we have control over. How can a non-causal relation, like essential metaphysical dependence, provide the kind of control necessary for our volitions to count as actions *of ours*? Even if our simple mental acts are caused, my claim is that they *count* as acts by virtue of being essentially grounded in our interactive mental properties. Since control is *intrinsic* to action, it must be *intrinsic* to those acts that they are within our control. Hence, for my view to be plausible control must be explained by the essential dependence relation.

This presents two challenges for my account. First, essential dependence—framed in terms of grounding—is too thin to adequately explain *how* simple mental acts might have their active character. Of course, I argued earlier that grounding relations back metaphysical *explanation* and so do not imply passivity for the grounded side of the relation. This objection, however, cuts deeper. Though I defended against the criticism that dependence implies passivity,

⁸⁷ Some evidence that we already implicitly accept something like these fundamental properties in both philosophical and folk language includes: (i.) that it is not uncommon philosophically to speak of the “reasons-responsive” mechanism of minds, which I have suggested correlate with my interactive properties though I hesitate to use the label because of reductive connotations, and (ii.) that it not uncommon in ordinary language to speak of “the will” rather than mere willings, where this seems to be indicate some underlying faculty of the mind.

I have not yet given an adequate account for how it explains *activity*. Second, essential metaphysical dependence does not adequately allow for control because it is *necessary* relation. That is, the ground entity entails the grounded entity necessarily and such an entailment undermines the open nature of deliberation. Both of these objections are plausible on their face; however, they fail for similar reasons. Explaining these reasons help to clarify what, exactly, my account aims to show.

The first objection is answered quickly, though perhaps in a manner unsatisfying to a skeptic. The active nature of simple mental acts is adequately explained in part by the intrinsic nature of the thing grounding my mental acts. As McCann (1998) says, we can explain why volitions are voluntary in the same way we explain water is wet—intrinsically. While my view is one step removed, because it appeals to the grounding relation, it still grounds those volitions in mental properties that are both fundamental within the domain of the mental and are interactive such that they constitute our agency. To paraphrase McCann, we can explain why those mental properties constitute our agency are active in the same way we can explain why water is wet. To cite interactive mental properties at all is already to give a kind of explanation, they *just are* what agency consists in and citing something that is *intrinsically* constitutive of agency *just is* to show that it is something under our control. Put differently, part of what makes this an “agency-first” account is my refusal to decompose agency into its “parts.” This is because, as I have shown, to do this obscures the very things we are trying to explain. Agency emerges *as such* and should be analyzed non-reductively. If so, then it is a mistake to think that we can provide an independent understanding of “activity” and then ask how agency meets this understanding. Activity and agency come in together and are in some sense inseparable.⁸⁸

⁸⁸ This point is, I believe, related to Thomas Reid’s famous claim that causation is primarily an *agential* concept that we then apply to the world at large. That is, according to Reid, we come to know what it means to cause something

To bolster this response, consider again an analogy to consciousness. It is no puzzle that we are aware of or have qualitative experiences of events that are grounded in consciousness/ Why not? In part because the properties of consciousness *just are* the unifying feature of minds like ours that explains the “qualitative awareness” of all the disparate events of which we are aware. In like manner, the interactive mental properties *just are* the unifying feature of minds like ours that explains the “activity” of all these disparate agential events in our minds. This is not yet to say that my account adequately explains the *mechanism* for active control, merely that the burden of proof is not quite as high as the objection initially proposes.

This response to the initial objection might seem too fast, but it is important to note that the dialectic is one in which I’ve already defended against the claim that my account entails passivity *and* my opponent as already granted the existence of mental properties underlying agency. Furthermore, they have already granted that these properties are fundamental within the domain of the mental. In this context, drawing the reader’s attention to the fundamental nature of those properties and how their existence already implies their active nature is both informative and sufficient to defuse the worry.

My critics might attempt to frame this worry more precisely. Instead of a general concern about the metaphysical dependence being too *thin* to explain active control they might worry that my account merely defers explanation. On my account, a simple mental event counts as an action because its essence metaphysically depends on mental properties that constitute agency. But metaphysical dependence, understood as grounding, is a necessary relation. So, we have the grounded entity whenever we have the ground, necessarily. In the case of action, this grounding

by virtue of *our experience* of causing things and then apply this concept widely (and perhaps mistakenly) to account for events in the world. I doubt Reid’s account of causation gives the full story, however, applying the same logic to the concept of “activity” or “active control” seems more plausible. Activity is, in the first instance, agential and we come to know what it means *by being agents*.

relation obtains between a *particular* simple mental event and the agential properties of mind. For example, suppose mental properties P_1 ground the essential features of volition V_1 at time t_1 . Now, at time t_2 the agent wills a new volition V_2 . We can ask, “what explains the essential character of V_2 ?” It cannot be P_1 because that entails V_1 not V_2 . So, the ground must have changed to P_2 to accommodate this shift.

If so, then it seems like my account merely defers explaining what gives us, as agents, our distinctive control over simple mental acts. The change in the ground is what *actually* provides a metaphysically robust explanation. Moreover, we might then ask about the nature of this change. If this change is causal then, in fact, causation *is* at the heart of what makes a simple mental event count as an action. After all, in this scenario causal changes, at least indirectly, determine *what essential character* the volition has.

This version of the objection clarifies *why* someone might think that a dependence account of action is inadequate for explaining agential control. It fails, however, because it misrepresents the scope of my account. A key feature of my account, which I discuss in more detail in section 7, is its compatibility with a causal account of action generation. Much like Ginet, I claim that the most plausible non-causal account of action should leave open the possibility that simple mental acts are caused, they are just not actions *in virtue* of anything causal. I argue that my account does this because the conditions for something's existence are different than the conditions that explain its essences, thus even if my simple mental acts are causally brought about they are only *acts* in virtue of their essential dependence on my agency.

Hence, *even if* changes in the ground properties are caused this does not matter for the volition expressing the essential character that it does. Though this causal explanation might explain the existence of such events, their essential qualities *including* the intrinsically active

control agents have over them is explained by their dependence on the agential properties of minds. As noted above, such agential properties are what *constitute* agency, and thus no further explanation is needed regarding why *they* count as something we control.⁸⁹

At this point, a critic might note that a *necessary* relation like metaphysical dependence undermines the openness of deliberation. That is, my account cannot adequately explain either contrastive questions (e.g. why did the agent choose this *rather than* that) or free action (e.g. could the agent will otherwise?). I do think that this concern merits further consideration, however, I reemphasize that my account is an account of what makes an event *count as an action* not what makes it count as a *free* action. These concerns track a worthwhile and interesting question for an account of free action, but it is not a question my argument needs to address.

My account, then, can explain control adequately because active control is an intrinsic essential feature of our simple mental acts and thus it is grounded in our *agency*, understood as non-reductively constituted by fundamental interactive mental properties. Our *agency* is paradigmatically (and essentially) something that *generates* control, thus though an independent analysis of what control *amounts to* might be desirable as a supplementary project, my account needs to go no further to explain the *mechanism* or *source* of control. Agency and control emerge together. This response, however, depends on the intelligibility of my earlier and oft repeated claim that my account is *compatible* with our actions being caused. It is time to examine this claim in detail.

⁸⁹ I also believe that my critic is relying on the controversial premise that all change is *causal* change. This claim, however, does not seem obvious to me. This would, however, require further argument; for now it suffices to point out that my account is compatible with the claim that the ground properties *causally* change.

7. Essential Metaphysical Dependence as Weak Non-Causalism

I follow Ginet (2002) in thinking that a non-causal theory of action should be compatible with causation. We should say that agents in a wholly causally structured world are *acting*, they are just not acting freely (p. 403). In Ginet's account this compatibility comes from the fact that though simple mental events are actions *by virtue of* their actish phenomenal quality this is consistent with them being causally produced.

Similarly, essential metaphysical dependence leaves space for agents in a causally structured world to still *act*. This is because my account grounds the *essence* of our simple mental acts in mental properties and the conditions that characterize *essence* are separable from conditions for existence. For example, the dog breed, "schnauzer," might have certain properties that obtain by virtue of its constitutive essence—such as, having four legs, being obedient, not shedding, having a beard, etc. That any schnauzers exist, however, is not contained in this essence. It is not *essential* to schnauzers that they exist, instead the breed's existence depends on many features outside of their essence—such as, the state of Holy Roman Empire in the 15th century, the practical need for rat-catching dogs in medieval Europe, the intentional actions of specific medieval breeders, etc.

Of course, considerations raised by Kripke (1972) regarding the necessity of origin might seem to suggest that for any *particular* entity its existence conditions *are* given by its essence (pp. 110-116). A skeptic could rhetorically ask, "isn't it the case that *this* particular schnauzer—named Scout, let's say—essentially depends on the conditions for its existence?" Scout could not have had different parents because were he to have had different parents he would be a different *particular* schnauzer. As such, the conditions for existential dependence and essential

dependence collapse together. It *is* part of Scout's essence that he was brought about *exactly* in the manner that he was.

This is a challenge for my view because if it is *possible* for it to be *essential* to our mental acts that they be caused (that is, for the conditions for their existence to be given *by* their essence), then it seems that in saying what makes a mental event an action I *must* make reference to their causal production, at least in those possible worlds wherein mental acts are causally produced. But, if this is true, then my account of action sometimes requires that it is *sufficient* to appeal to a certain causal structure to characterize a mental action even though it is not *necessary*. To admit this, however, raises questions about whether or not my account is needlessly complex, if an event-causal story *can* explain why this mental event counts as an action then perhaps we need only appeal to causation.

This raises many important questions. The first distinction to note is that there is a difference between saying "that something essentially exists" and "that if something exists it essentially exists in this way." Even in the case of Scout, it is not true that he *essentially* exists only that since he exists the conditions for his existence are essential to him. To have Scout exist in another possible world we must replicate those same conditions for existence *in that world*. Is it right, however, to conceive of this as an example of Scout's essence? We've already distinguished between modal conditions and essential conditions, these two conditions come apart in interesting ways. Might it be the case that it is *necessary* for Scout to have the parents he does but not essential? Alternatively, perhaps necessity of origin is part of the consequential essence of Scout, without being itself part of its constitutive essence. I think both of these suggestions might allow wiggle room to make the distinction between existential and essential in

the case of a *particular instance* of some kind, however, below I consider an alternative response that is even more direct.

Note that the reason *this schnauzer's* origin is essential is in part because in rigidly designating *this schnauzer* we name those things that *actually constitute* it. Whatever else we might say about schnauzers they are constituted by matter essentially and thus *this schnauzer* is constituted by *this* bit of matter essentially, hence its parents *who are determinative of this bit of matter* are essential to *this* particular schnauzer. Is there an analogous move in action? It does seem that there is a kind of necessity of belonging, it seems senseless to ask “could someone else have willed my willings?” It is necessary that *this* willing belong to *me*.

Does that mean this willings' manner of coming to exist is essential such that I cannot distinguish between the existential and essential manners of dependence? Perhaps not. What am I pointing to when I indicate *this* volition? In this case I “point” to something *mental*, not a bit of matter as such, rather what *this* volition has as its essential constitutive features is the *rational* content incorporated into the volitions by its grounding in interactive properties of mind. This point might seem to presuppose a controversial picture of mind, but note that all this requires is the falsity of naïve mind/brain identity theories, it is still compatible with many other views of mind.

Let me give some examples. My claim is that even in the case of particular volitions we can separate the conditions of existential dependence from the conditions for essential dependence such that a *particular* simple mental act would still be an act of mine by virtue of essential metaphysical dependence *while its existence* is explained by other factors, like causal forces. This is so my view can be compatible with the plausible thesis that in a causally structured world we would still *act* (unfreely), however, events would not *count* as actions by

virtue of anything causal. To establish this, the key question is whether or not essential dependence can adequately explain how particular actions belong to an agent independent from existential considerations *even in* those worlds that are causally structured. I argue for this by showing how all of the conditions that are physically required for causal existence of a simple mental event might be met *without* the event counting as an action of the agent's, thus providing a counter-example. Consider the following case:

Equivalent Exchange: Suppose Non-reductive materialism and functional-realization accounts are true. Mental properties and simple mental events, like volitions, are caused by brain-states which functionally realize mental states. This means that different physical bases (e.g brain-states, silicon-states, distributed neural network states, etc) can realize the same token mental state. An agent, Diane, has a device implanted in her brain by Martians as a baby. This device functions to help realize volitions through normal causal processes. As such, when the device realizes the volition the essential features of that volition (e.g. being constituted by the reason for which it is done, having a certain phenomenal quality, etc.) are grounded in interactive mental properties, which themselves are functionally realized by a different part of the brain. This device, however, has a secondary function. Whenever Diane thinks about moving to Denver, Colorado the device *instantly* transmutes her entire brain into a silicon based brain. The causal structures that would normally produce brain-states remain unchanged, only the nature of the physical base has changed. Diane wills to move to Denver.

How should we understand this case? It seems that certain conditions for the *existence* of Diane's volition have changed but it is still essentially grounded in the same mental property. After all, that mental property can be realized by different physical basis. This, however, does not yet show that my *particular* volitions are actions by virtue of their essential grounding alone, at best it shows that the existence conditions for action are generic but still essentially required to count as an action of mine. Here is a similar case:

Equivalent Exchange*: Everything is the same as *Equivalent Exchange*, except the Martian device only replaces those physical substrates that underlie interactive mental properties with a silicone based substance. Diane wills to move to Denver.

Again, it seems two things remain the same. First, the causal chain that leads to these physical substrates existing in the manner they do is structurally the same. Second, the mental states that are realized by this substrate manifest the same dependence relationships, such that the volition to move to Denver is grounded in the interactive properties of mind. The *targeted* change in the substrate does not affect the causal relations that uphold the existence of their realized mental types any more than the whole brain switch. Again, the point here is that the conditions for the *existence* of Diane's volition to move to Denver are not necessary for them to be an action of hers. Of course, in one sense they are necessary insofar as *some* physical base is necessary but this does not uniquely select the mental event as an action. Instead, its counting as an action depends on the essential characteristics of *that* volition being grounded in *that* mental property, neither of which make reference to *that* physical base (since it is compatible with many physical bases).

The skeptic should still protest that while it is not necessary for a simple mental act to have its essential character it is nevertheless sufficient. Surely if we know that a particular kind of physical structure exists that can functionally realize a particular sort of mental property, and its dependent mental events, then we also know that there is an action. Hence, the underlying physical/causal structure is sufficient for characterizing when an event counts as a mental act.

Ginet (1990) considers a similar objection to his view (pp. 10-12). If volitions are caused by neural structures in our brain then we should be able to explain what makes them an action by reference to those structures. Ginet has two responses, first he notes that it is not *conceptually required* for our simple mental acts to have an event causal structure. That is, the concept of a mental act does not require an appeal to causation regardless of whether or not it is produced by a causally structured neural process (or is even identical to that process). So, when we mentally

say the French word *peu* as opposed to just having the French word *peu* come to our minds, we must recognize the difference because of something *intrinsic* to the event. There is something in the concept of a mental act that sets it apart from a mere mental happening and we recognize it in the event. Whatever this indicator is that is intrinsic to the event it cannot be the fact that it is caused.

This response, as stated, is not adequate. For, the causalist might respond that while it is not conceptually part of our phenomenal concept of action it might be the case that actions have causal components that are inaccessible to consciousness. This response is, in some ways, analogous to the move Papineau (2002) makes when discussing the hard question of consciousness.⁹⁰ A causalist about action could make a similar move against Ginet and argue that there are distinct *phenomenal* concepts by which we as conscious agents are *aware* of our actions but this is merely an epistemic criterion that need not imply any deep metaphysical realities.

Ginet's (1990) second response is given in his discussion of agent-causal views but it also applies to the event-causal views of neural-structures. If our phenomenal experience of acting, Ginet says, is merely a *sign* of an underlying causal structure then we need some explanation for what makes the difference between veridical and non-veridical phenomenal experiences. Though he does not say it directly here I believe Ginet is gesturing towards the problems of causal deviance, he seems to think that extrinsic relations always have the possibility of obtaining deviantly. Thus, we need some explanation for what makes the difference between mental events and mental acts in the two cases. Ginet argues that since there is no phenomenally accessible

⁹⁰ Briefly, Papineau (2002) argues that materialists should be *conceptual* dualists and ontological monists about mental properties. While we can refer to mental properties *as* material, it is also possible to refer to them as phenomenal in virtue of how they feel. This does not mean they are non-physical, rather, it means that we have two very different ways of think about the same physical property. Papineau argues that this difference in conceptual reference is a result of the self-referential nature of conscious thought.

difference on the causalist view then we really have no idea what the agent-cause (or event-cause) really is with respect to action. It is far more parsimonious to just appeal to the phenomenal quality as making the difference between an act and a mere mental event.

Whatever the merits of this response for defending Ginet's view, it will not be adequate to defend mine. As noted above essential metaphysical dependence is a *necessary* connection and thus it cannot fall victim to deviant chains. This is a boon for explaining how the essential phenomenal character of our volitions might be a veridical indicator of the underlying extrinsic metaphysical relation, however, once one grants this necessary connection it becomes harder to separate essential conditions from the causal forces that underlie the existence of those acts in causally structured worlds.

How then can I respond to the above criticism? Consider this final case:

Equivalent Alien Exchange: Everything is the same as in *Equivalent Exchange** except when the device targets the physical base for Diane's interactive mental properties it connects them instantly (via some sort of Martian wifi) to the physical base of another agent, Clayton, who is *also* deciding to move to Denver. It is the interactive mental properties that are realized by the alternate physical base that figure in Diane's decision to move to Denver, and vice versa.

What are we to make of this case? The previous examples make clear that differences between physical basis do not matter with respect to the mental event that is realized, so the *mere* fact that a different physical base is realizing the interactive mental property does not change anything.

Furthermore, the causal connections that lead to these physical states are going to be, I presume, structurally isomorphic—both of them are brain states that are analogously produced. I contend, however, that the volition that figures in Diane's deliberation is *not* an action of hers, even though the causal conditions for its *existence* in principle pose no problem for it being an action of Diane's. Instead, it is because the *volition* is itself disconnected from any mental property of

Diane's—it is Clayton's mental property. It seems like the right account of the case is that the mental event does not count as an action of Diane's because it is not *grounded* in her agency and this is true regardless of the causal conditions for its existence.

So, I conclude, it is possible that a condition for the *existence* of a given simple mental act is some neural-event or brain state that causally explains the *existence* of the particular simple mental event. What makes this simple mental event essentially an *act*, however, is the grounding relation between the *agent* and the simple mental event not the causal structure that explains its existence.

8. Conclusion

Action as essential metaphysical dependence combines the most desirable elements from standard causal accounts, non-standard agent-causal accounts, and traditional non-causal accounts. I agree with standard causalists that to make actions explicable in a world that is otherwise causally structured we should appeal to an extrinsic relation; thus, avoiding the methodological dead-end of dialectical stalemates. The non-causalists, however, are right to think that the extrinsic relation of *causation* is unsuited to this task. Not only do causal relations fail to account for what is distinctive about human agency, they also obscure agency itself. This failure to adequately keep the agent in view is endemic to most standard accounts of action. One desirable feature of agent-causal theories is how they link actions directly to the agent. My account, by defining simple mental acts *in terms* of their dependence on fundamental mental properties that constitute agency meets this final desideratum making it a true “agency-first” account. Moreover, this manner of explaining actions clarifies and resolves the intractable problems of “activity” and “control” that follow all theories of action. This already does much to recommend this account as viable, however, there is one final objection that must be addressed.

I analyze dependence in terms of “grounding” or a grounding-like notion. As I noted earlier, grounding is traditionally understood as transitive. If so, then there is a serious objection waiting for my account. If what makes my simple mental event count as an action is that it is grounded in fundamental mental properties, but those mental properties are (in turn) grounded in some further set of natural properties, then it seems my actions are *also* explained by virtue of being grounded in these further properties. Transitivity means that the agent, once again, disappears. Agency does not *really* explain what makes an event count as an action. This objection is challenging, but I believe there is a response. In the final chapter—chapter 4—I argue that the concept of “emergence” can explain how there is enough of an ontological “gap” to avoid the transitivity objection. Moreover, such an emergent account of mental properties fits well with the non-reductive picture of agency I propose.

CHAPTER 4: EMERGENCE AND THE TRANSITIVITY OBJECTION

I claim that what makes a simple mental event count as an action is that it essentially metaphysically depends on the interactive mental properties that constitute agency—that is, it is *grounded* in agential minds. This account of action is vulnerable to the “transitivity objection.” Grounding is, typically, understood as a *transitive* relation. If A grounds B and B grounds C, then A grounds C. Since I analyze “dependence” in terms of grounding, and since grounding is transitive, it seems like my simple mental events are not *only* grounded in our agency but also grounded in the underlying physical features that ground agency itself. This is a problem for two reasons: (i.) if the dependence relation is what *makes* an event count as an action then my account becomes too permissive, what explains actions can be the same underlying dependence relations that explain non-actions. (ii.) If actions are explained by being grounded in properties *outside* of our agency, then the agent does not explain the action in any meaningful way, the agent disappears from my account of action.

This objection is challenging, but it can be met by allowing a “gap,” which breaks the chain of transitivity. The central difficulty in this response is making the “gap” a plausible feature of the world rather than just an *ad hoc* response to an objection. I argue that the concept of “*emergence*” provides just such a solution. If agential-minds are emergent features of the world, then I argue there is enough of a gap between the mental and the physical to allow our actions to properly depend on *us* not on the underlying laws that *ground* us. Moreover, since emergence still posits *some* sort of dependence between the emergent property and its “base”, an emergent account would be compatible with our larger naturalist conception of the world. Indeed, I hope to show that an emergent view is far from “radical” and may be *supported* by our best science.

In §1, I argue for this, first, by introducing the concept of emergence in general. I briefly describe its intellectual history and connect it to various contemporary projects in metaphysics. Second, I examine more closely the conceptual debates regarding emergence. In particular, I focus on debates between unitary accounts of emergence, which describe emergence as a single unified concept, and pluralist accounts of emergence, which describe it as having several distinct *types*. In §2, I conclude that a synthetic understanding of the concept is most justified, wherein the unitary accounts explain what emergence *amounts* to while the pluralist accounts explain *how* it amounts to this. In the second half of the chapter, starting in §3, I turn to apply the concept to my account of action, reveal what type or “degree” of emergence is needed for my account. Finally, in §4, I consider the plausibility of emergence by briefly canvassing its use in various sciences, I pay special attention to whether instances of “strong” emergence can be justifiably posited given our best science. I conclude that not only can emergence solve the transitivity problem, but it is open given our best science that such ontological emergence exists in our world.

1. Emergence—History and Conceptual Development

An emergent property is, roughly, a property that is (in some sense) *dependent* on the base properties that constitute or support it, while at the same time (in some sense) *autonomous* from those properties (Corradini & O’Connor 2010, p. 3).⁹¹ Emergence was coined, as a concept, to make sense of the seemingly independent nature of the special sciences. This tradition of ‘British Emergentism,’ which started in the mid-nineteenth century and lasted well into the early 20th century, began with John Stuart Mill’s *System of Logic* in 1843 and found its most mature

⁹¹ These two hallmarks of emergence are often described in passing (Bedau 1997, p. 374; 2010, p.47; Wilson 2015, p. 346), but more often they are, as Wilson (2015) notes, embedded in particular accounts of *what* those senses of dependence and autonomy are (pp. 346-347, fn. 3).

articulation in C.D. Broad's *The Mind and Its Place in Nature* in 1925 (McLaughlin 1992, p. 49). The British Emergentists were concerned with understanding the "levels" of reality as demarcated by the special sciences. By their lights, though all entities were composed of matter and governed by properties described by physics, some "higher" level entities described by the special sciences possess causal powers that emerge because of those kind material structures that they possess.

As McLaughlin (1992) notes, these arguments were motivated by the failure of reductive projects between the chemical and physical levels of reality in the 18th and 19th centuries. This made plausible the claim that there may be forces, and thus force laws, that are configurational, which is to say that the law *depends* on the configuration of the lower-level entities that constitute the higher-level entity (pp. 52-55). This *specific* project was rejected after the successful explanation of chemical "forces" in terms of quantum mechanical understandings of electro-magnetism (McLaughlin 1992, p.54).⁹²

This failure made emergence fall out of favor for some years, but as non-reductive approaches in philosophy of mind experienced a resurgence of interests in the mid-20th century (see, Putnam 1963, 1967; Davidson 1970; Kripke 1972) emergence, of varying sorts, started to gain ground once again. In addition, arguments against the reduction of the special sciences to physics (Boyd 1980; Fodor 1974; Putnam 1973) made the concept of emergence an attractive possibility for non-reductive materialists. In the sciences suspicion of non-reductive projects continued, bolstered by the continuing success of molecular explanations in biology and quantum mechanical explanations of chemistry. However, by the 1980 the advent of complexity

⁹² McLaughlin (1992) points out that the last major work by the British Emergentists was by C.D. Broad in 1925 was immediately followed by the quantum mechanical revolution that spelled the downfall for the British Emergentists project as it was conceived by its proponents (p. 57), further this point has been recognized by earlier philosophers of science such as Nagel (1952).

studies, information science, and computer science, had once again shifted the weight of discussion. To be sure, reductive methodology is still widely used within scientific discourse, but emergent theories are once again plausible within certain domains.

This large-scale development of emergent concepts in metaphysics has led to a proliferation of theories. As noted above, all theories of emergence call for the emergent property to be (in some sense) dependent on the base properties that constitute or support it, while at the same time being (in some sense) autonomous from those properties. It is widely agreed (Wilson 2015; Taylor 2015a; Kim 2006; O'Connor 1994), however, that explicating these two features of emergent properties is a messy project. As such, many recent articles on emergence spend a significant amount of time taxonomizing the complex theoretical terrain. My account is no exception. I consider four general types of emergent autonomy that help categorize these diverse theories—strong, weak, ontological, and epistemic. Before I discuss the four general types of emergence, I introduce yet another distinction this time between research programs that philosophers pursue in characterizing emergence.

I distinguish between the conceptual project and the empirical project. The conceptual project aims to clarify what we mean by emergence, while remaining agnostic about whether any emergent entities actually exist. By contrast, the empirical project aims to describe what emergent properties or substances, in fact, exist. Most philosophers in the emergence literature are engaged in both projects to some degree. Depending on the theories, the order of priority between the two projects changes. Some philosophers clarify what they take to be the most rationally defensible concept of emergence and then use the newly clarified concept to determine what counts as existing emergent phenomena (Dennett 1991; O'Connor 1994; O'Connor & Wong 2005; Taylor 2015a). By contrast, others start with the empirical phenomena, for example

uncontroversial cases of emergence in scientific practice, and use these considerations to clarify the concept (e.g. Batterman 2011; Bedau 1997, 2010; Castellani 2002; Kronz & Tiehen 2002; Wilson 2013; Gillett 2016). The conceptual project can be further divided between those who argue for a plurality of concepts (Bedau 2010) and those who argue that there is a single unified concept of emergence (Taylor 2015a).

In this section, I primarily take up the conceptual project, although I inform this project by reference to various examples of emergence that any adequate theory should take into consideration. Specifically, I argue that we should be pluralists about the concept of emergence. A pluralist about emergence believes that adequately explaining emergent phenomenon requires at least two (perhaps several) distinct concepts that share a family resemblance. I take up the standard distinctions between nominal, weak, and strong emergence. But I also argue for variety between the types of autonomy that the emergent property might possess: epistemic, causal, or structural. I defend these claims in the next section against Taylor's (2015b, 2015a) recent arguments for a unified conception of emergence defined in terms of explanatory failure. I conclude that while Taylor does identify a property that all forms of emergence possess, it is not sufficient to define the concept solely in those terms. Now that the type of project I am pursuing is clearly in view, I turn to taxonomize emergence by examining the four emergent considerations—strong, weak, epistemic, and ontological.

When characterizing emergence, theorists first distinguish between whether the emergence under consideration is “strong” or “weak.” Roughly, the strong/weak distinction tracks the *degree* of autonomy that an emergent property has from its base. A strongly emergent property is more autonomous from its base, a weakly emergent property less so. This autonomy is then characterized in either ontological or epistemic terms. These two sets of considerations

cut orthogonally across each other. For example, Chalmers (2006) suggests that the strong/weak distinction tracks whether the emergent property is un-deducible from our knowledge of the physical base as opposed to merely surprising or unexpected—this consideration is epistemic (pp. 244-245). In contrast, Wilson (2015) argues that the strong/weak distinction tracks whether the emergent property has at least one more power than its base, or if it has less powers than its base—this consideration is ontological (p. 362).

This suggests that the primary consideration for characterizing an emergence property is the autonomous side of emergence. That is, philosophers *first* attempt to distinguish both the degree (strong/weak) and the characterization (ontological/epistemic) of the autonomy conferring feature before considering other issues. This is a sensible method but it should not obscure the *other* side of emergence—dependence. In what sense does an emergent entity *depend* on its base and how is this dependence characterized? Obviously these two lines of investigation overlap, different explanations of dependence are compatible with different considerations of autonomy. In the interest of tracking the dialectic accurately, however, it is useful to remember that characterizing autonomy is vital for establishing the type of dependence prior to characterizing *how* such an autonomous feature might depend on its base.

With these general considerations in mind, I briefly note in Table 1 some ways that emergent autonomy has been characterized. Given this large but still partial list of distinctions, we should ask if there is any unifying feature that allows us to simplify the conceptual space. I end this section by considering two such attempts, which inspire the conceptual debate I raise in the next section—Wilson’s (2015) “powers” account and Taylor’s (2015a) “explanatory failure” account. Wilson (2015) argues that the myriad of differences among accounts of emergence are, in fact, superficial (pp. 347-349).

Table 1. Varieties of Emergence.

	Strong Emergence	Weak Emergence
Ontological Characterization	<p><i>Novel Powers-</i> The emergent entity or property has a new causal capacity not found in its base properties or by a mere summation of the powers of these base properties.</p> <p><i>Failure of Realization-</i> The emergent entity is not functionally reducible to the set base properties as its realizer.</p> <p><i>Non- linearity (strong)-</i> The emergent entity has features that cannot be additively derived from the base properties. The explanation for this failure is a novel capacity interacting with the base properties.</p> <p>Etc. . .</p>	<p><i>Proper-subset of Powers-</i> The emergent entity possesses a non-empty proper subset of the token powers of the base properties.</p> <p><i>Degrees of Freedom (limit)-</i> The emergent entity requires less independent parameters to characterizes its states than its base properties.</p> <p><i>Non-linearity (weak)-</i> The emergent entity has features that cannot be additively derived from the base properties. The explanation for this failure involves the complex interactions of microlevel entities in the base properties that can only be identified at the macrolevel.</p> <p>Etc. . .</p>
Epistemic Characterization	<p><i>Failure of Conceptual Entailment-</i> The emergent entity does not logically supervene on the its base. This means that an idealized conceiver would, in contemplating the base entity and the emergent entity, be able to conceptually conceive of one without the other.</p> <p>Etc. . .</p>	<p><i>Explanatory Incompressibility-</i> The emergent entity can be derived from its base properties, but only by iterating the systems' microdynamic while taking the initial starting conditions as inputs. That is to say, the emergent entity is derivable only by "simulation."</p> <p>Etc. . .</p>

Wilson claims that, given what it would take for an emergent entity to be metaphysically distinct from its base, there are two and only two schemas for characterizing emergence.

Moreover, both characterizations involve constraints imposed on the powers of the emergent property with respect to the base. These two strategies track the “strong” vs. “weak” distinction that other theories have indicated, but they do so in a more principled manner:

Strong Emergence: Token higher-level feature *S* is strongly metaphysically emergent from token lower-level feature *P*, on a given occasion, just in case (i) *S* synchronically depends on *P* on that occasion; and (ii) *S* has at least one token power not identical with any token power of *P* on that occasion. (p. 362)

Weak emergence: Token higher-level feature *S* is weakly metaphysically emergent from token lower-level feature *P* on a given occasion just in case (i) *S* synchronically depends on *P* on that occasion; and (ii) *S* has a non-empty proper subset of the token powers had by *P*, on that occasion. (p. 362)

Wilson (2015) means “powers” to be understood in as metaphysically neutral a manner as possible. As she puts it, “talk of powers is simply shorthand for talk of what causal contributions possession of a given feature makes (or can make, relative to the same laws of nature) to an entity’s bring about an effective, when in certain circumstances” (p. 354). This is meant to track the uncontroversial thesis that the way entities are contributes to what entities can do. Given this frame, the *Strong Emergence* thesis captures well the intuitive claim that what matters most for an emergent entity to be *strongly* autonomous is that it makes contributions to how the world “unfolds” that are not reducible to the contributions of any of its parts.

The characterization of *Weak Emergence* requires more explanation. The proper subset strategy was first posited and developed by Wilson (2011, 1999) to solve puzzles regarding mental causation, however, it is useful as a general characterization of weak emergence.⁹³ It is motivated, primarily, by cases of multiple realizability. The idea is that powers of a multiply

⁹³ A strategy similar to this has, predominately, been associated with Shoemaker’s (2007, 2001) work on realizability. Wilson’s (1999) earliest account, however, predates these later publications.

realized type are those in the intersection of the sets of powers of its realizing features, and hence these powers are part of a proper subset at the token level (p. 358). This schema ensures autonomy in two ways, first it ensures ontological autonomy by simple non-identity. If the higher-level emergent entity has only a proper subset of its base's powers then it is distinct from its base by Leibniz's law. Second, it ensures causal autonomy by having a distinctive "power profile" (i.e. a set, collection, or plurality of powers), rather than a distinctive power (p. 360). One way to see the plausibility of this is to consider how a proper-subset view supports a counterfactual or "difference-making" account of causal explanation. For example, suppose that some mental state, *M*, is multiply realizable. Further, suppose, *M*, is the state of feeling hungry, which causes an effect, *E*, of reaching for an apple. In this instance, *M* was realized by physical state *P*, but suppose it had instead been realized by *P**. Would *E* still have occurred? It seems the answer is, yes, because the differences between *P* and *P** do not matter for *E*, all that matters is the powers associated with *M*. That is, all that matters is *M*'s *distinctive power profile*, which contains just the powers crucial for *E*. This gives us, Wilson (2015) argues, a principled reason for taking *M* to be efficacious with respect to *E* in a way distinct from *P* or *P** (p. 361).⁹⁴

Another, additional reasons to think that *Weak Emergence* still can guarantee a certain kind of causal autonomy is that distinctive power profiles (at least sometimes) are governed by (or ground) distinctive systems of laws. It is plausible that for certain special sciences, the laws they posit track the powers of the emergent type *S* that has a proper subset of the powers of its physical realizer, *P*. Moreover, it is plausible to think that systems of laws track causal features in nature. Hence, *S*'s distinctive power profile may indicate a distinctive causal feature in nature.

⁹⁴ As Wilson (2015) notes, part of why this emergent schema is considered a version of weak emergence is because it is compatible with *E* also being caused by *P* or *P**. The thought is merely that difference-making considerations give a principled reasons to think that *M* is distinctively efficacious with respect to *E*, insofar as *M*'s power profile tracks those powers that are relevant to the production of *E* across counterfactual situations (p. 361).

Such law-like causal stories might overlap, as in the overlap of the subset of powers shared by *S* and *P*. But, if these causal features are, as a whole, different then there is a principled basis for taking *S* to be distinctively efficacious with respect to its effects insofar as its powers are part of a difference system of laws than *P*.⁹⁵

Hence, the two schema's—*Strong Emergence* and *Weak Emergence*—characterize emergent accounts by specifying the different ways powers in an emergent entity may differ from its base. Moreover, Wilson (2015) argues that, given the difficulties of maintaining the independence of higher-level causation, these two schema provide the *only* principled basis for explaining the metaphysical emergence of higher-level broadly scientific entities by ensuring genuine autonomy. Strong autonomy, by virtue have having *more* powers than the base entity; and weak autonomy, by having *less* powers than the base entity (p. 363). Wilson (2015) thinks that the other, more specific accounts of emergence considered above all conform to one of these two schema and argues for this in detail (pp. 363-388). She also argues that *epistemic* accounts of emergence are either *backed* by some sort of ontological account (e.g. Chalmer's 1996, 2006, understanding of *failure of conceptual entailment* is meant to establish the *strong emergence* scheme suggested by Wilson.) or are not a genuine account of emergence but rather reductive physicalism offering a purely epistemic error theory for explanatory gaps. Of course, Wilson recognizes that the motivations for these reductive epistemic accounts tracks broader motivations for holding “genuine” emergent views (i.e. our failures in explaining higher-level entities). But, these accounts *assume* ontological and causal reduction and locate the appearance of autonomy in a mere failure of cognitive connection; hence, by Wilson's lights, they ought not be considered emergent accounts.

⁹⁵ Wilson (2010) further defends this claim in the context of a “degree of freedom” account of non-reductive materialism.

This general schema of emergences which conceives of it as being centrally a *metaphysical* concept capturing broad dependence and power based ontological and causal autonomy is a good foil to Taylor's (2015a) alternative schema for bring conceptual order to the house of emergence. Taylor (2015a) argues that emergence is best understood as a unitary concept that tracks explanatory failure of a certain kind. The motivation for sort of account is twofold, first, it provides a unified concept that is flexible enough to be used by both scientists and philosophers, thus allowing productive discussions about emergence without merely talking past each other. Second, it solves a problem that Taylor (2015b) thinks is endemic to all emergent theories, the *collapse problem*. In the rest of this section I present Taylor's (2015a) account and explain the primary differences between it and Wilson's account. The *collapse problem* I save till the next section, wherein I evaluate which of these conceptual accounts of emergence is the most successful.

Taylor is motivated in part by a wider sort of conceptual unclarity. Taylor (2015a), like Wilson, is also concerned with the messiness of philosophical theorizing, but by her lights an equally important challenge is how the concept of emergence can be effectively deployed *across* scientific and philosophical discourse (p. 654). As she notes, many philosophers think that the difference between emergence as discussed in the sciences and emergence in philosophy is so great that they are different concepts (pp. 656-657).⁹⁶ Taylor (2015a) disagrees and claims that much scientific theorizing about emergences maps onto philosophical debates, however, the attendant philosophical unclarity obscures this fact (p. 658). To address this unclarity for the

⁹⁶ Taylor (2015a) cites Chalmers (2006, pp. 1-2 [sic. pp. 244-245 in my document]) and Stephan (2006, p. 485). In both examples, it is worth noting, the authors take the "scientific" concept of emergence to be something similar to weak emergence as spoken of by philosophers. In contrast, Gillett (2016) argues that not only are the scientific and philosophical concepts of emergence distinct, but this distinction cuts across the strong vs weak divide. That is, even philosophical concepts of "weak" emergences are still inadequate for explaining emergence as used in the sciences.

sake of improving both scientific and philosophical discourse Taylor (2015a) proposes using the method of “rational reconstruction” to more precisely define emergences *as used* in scientific and philosophical practice (p. 658).

Rational reconstruction involves taking a vague concept and making it more precise by attending to its *use* within various domains. Such a method aims to unify the concept by taking into account, insofar as is justifiable, the various uses of the concept. This project is not, Taylor (2015a) argues, *determined* by the original use; revision of the concept is possible (p. 658). Such revisions follow standard theoretical desideratum such as simplicity, fruitfulness, and accuracy. These considerations are then weighed against the primary aim of unifying discourse and preventing equivocation. Taylor (2015a) thinks that once this method is applied to emergence, we find that all of the significant discourse can be subsumed under an ecumenical conception of emergences as a kind of *explanatory failure* (pp. 658-659).

According to Taylor (2015a), emergence is best captured by the following explication:

(EM) Given components A, B, C . . . n arranged in relation r into a whole, and an observer O, property x of the whole is emergent for O iff there is no scientific explanation available to O of the fact that the following regularity obtains of natural necessity: *Whenever components A, B, C ... n are combined in relation r, the resulting whole instantiates property x.* (p. 659)

A few of these terms require clarification. First, in this explication “components” should be understood as merely a placeholder for whatever “bears” the base properties (or micro-level properties, as Taylor prefers to say) and “whole” should be understood as a placeholder for whatever bears the purportedly emergent properties (or macro-level properties). Second, the “observer” is meant to indicate *any* observer or *any* number of observers. Third, and finally, “explanation” should be understood very broadly. Though Taylor (2015a) considers it “scientific” this merely indicates that whatever form of explanation should be salient to our

understanding of empirical reality, insofar as this is the relevant domain of questions regarding emergence (p. 660). Within these broad boundaries the concept of explanation can encompass many different types, such as, mechanistic explanation, statistical explanation, deductive explanation, computational explanation, etc.

Taylor's (2015a) understanding of emergence, then, has several interesting features. First, it follows that *nothing* is emergent *per se* but only emergent relative to an observer with respect to a particular domain of explanation. This implies that a property's status as "emergent" can change depending on the observer and the domain of explanation. For example, the properties of chemical compounds are emergent with respect to mechanistic explanation, but not with respect to statistical quantum mechanical explanations. In like manner, a property might be emergent to actual human observers but not to an idealized observer. It also implies that emergence is *prevalent* and *cheap*, a point I will return to below. Second, it follows from Taylor's account that we cannot discern anything about *how* an emergent property obtains from the mere fact that it is emergent. Sometimes, Taylor (2015a) notes, a property might be emergent for metaphysical reasons at other times it might be emergent for epistemic reasons, but in both cases the presence of the emergent property itself only indicates a certain failure of explanation with respect to a certain (group of) observer(s) (p. 660).

Taylor (2015a) thinks that this conception is unifying in a *justifiable* manner because there is an important conceptual connection between emergence and explanation, which is present even in accounts that are not explicitly formulated in explanatory terms (p. 660). For example, given the wideness of her concept of explanation, failure of conceptual entailment, associated with Chalmers (1996), is a sort of explanatory failure. In like manner, a more ontological account, like O'Connor's (2000; O'Connor & Wong 2005), can be easily re-

described in explanatory terms, a point evident from O'Connor's own discussion. For example, O'Connor (2000) speaks of emergence as a natural, non-structural property possessed by systems that obtain a certain level of complexity, which exerts downward causation (p. 111). But, he goes on to explicate a "non-structural property" as a property with powers that are not *explainable* by reference to a mere summation of the powers of its micro-level constituents (pp. 111-112; see also p. 109). Likewise, he motivates accepting conceptually the distinctness of macro-level properties and their unique causal powers by discussing the potential failures of lower-level laws to *explain* all the physical processes of the system under consideration (p. 113). In both cases, the concept of emergence used by philosophers can be made compatible with EM without doing violence to their distinctive explanations for *how* the given property is emergent (presumably metaphysical reasons for O'Connor's and perhaps conceptual reasons for Chalmers).

Furthermore, Taylor (2015a) argues that EM is also preferable because its flexible account of explanation and shifting definition of an "observer" encompasses a wide variety of domains in both philosophy and sciences (p. 660). This allows it to meet the two primary aims of rational reconstruction—unifying discourse and preventing equivocation—better than alternative theories. This very flexibility might be taken as flaw in the theory. For example, if an observer has mistaken or little scientific knowledge compared to another observer then it follows that *more* entities would count as emergent for the first observer. Taylor (2015a) recognizes this but thinks that the prevalence and "cheapness" of emergences is a benefit of her view (p. 661). Explicating emergence in terms of EM reframes the discussion from *why* and *whither* an entity is emergent to *how* and *for whom* an entity is emergent. This shift, then, gives theorists *more* resources to consider what types of emergences are *important* and, importantly, *why* we consider them important.

For example, the emergence of an entity because of a failure of explanation that *depends* on the scientific ignorance of the observer has very little, if any, philosophical interest *qua* emergence. EM can explain this lack of importance *in terms* of the narrowness of its two main variables (i.e. the type(s) of observer(s) and the kind of explanation). In this example, the variables are greatly constrained. After all, the type of explanation at stake is not scientific explanation *as such* but rather the flawed *conception* of scientific explanation idiosyncratically held by a particular observer. In like manner, the observer in questions is a single, non-ideal, and likewise idiosyncratic person. In contrast, interesting cases of emergences can be explained by their *wide* construal of EM's variables. For example, strong metaphysical emergence would involve, plausibly, *all* observers and a robust and universally applicable type of scientific explanation. Taylor concludes that since EM does not *require* that all cases of emergences have equal importance *and* it gives us the resources to explain this differential importance in a principled way; hence, the fact that it makes emergence widespread is not a problem for the theory.

Wilson's (2015) power-based account and Taylor's (2015a) explanation-based account are both attempts to impose *conceptual unity* on the discourse about emergence. In Wilson's case this unity is *metaphysical*; in Taylor's case it is merely *explanatory* and therefore *epistemic*. Is one to be preferred? Taylor (2015b) has gone on to motivate her account by raising a general problem for all metaphysical accounts of emergence, which she calls the "collapse problem" (pp.732-733). She argues that because no alternative accounts can adequately address this problem her theory of emergence as explanatory failure is the only option. In the next section, I consider the arguments surrounding the collapse problem as a way to evaluate the power-based and explanation-based strategies of conceptual unification. I conclude that while metaphysical

accounts *can* survive and respond to the collapse problem, the *other* desideratum that Taylor (2015a) offers in her positive account requires that we take the *epistemic* construal seriously. I end this section by offering a hybrid view which takes a certain type of explanatory failure to be indicative of metaphysical emergence. This, then, allows us to return to apply my concept of emergence to the case action as essential metaphysical dependence and thus address the transitivity objection.

2. The Collapse Problem and a Hybrid Account of Emergence

Taylor (2015b) argues that all going versions of emergence face a generalized collapse problem. This problem arises because emergent accounts presuppose a distinction between the micro-level (base properties) and the macro-level (emergent properties), but for any instance of emergence there are properties that on the face of it belong to the *micro-level* yet if included at that level the purported emergent entity fails to be sufficiently autonomous (p. 732). Taylor (2015b) calls these properties, “collapse-inducing properties”, because once included in the micro-level the emergent macro-property effectively “collapses” down to that level (p. 732). Moreover, there seems to be no non-arbitrary way to exclude these collapse-inducing properties. Taylor (2015b) claims that this problem suggests that emergence tracks relationships between *arbitrary* properties, rather than distinctions in levels of nature (p. 733).

The method for generating collapse objections is simple: for any micro/macro distinction used in an account of emergence, find a micro-level property that crosses the line marking emergent autonomy *yet*, according to the original micro/macro distinction, is still solidly part of the micro-level (p. 744). This makes the problem quite broad. Here are two examples (one of which Taylor also discusses) taken from the views presented above:

(Ex. 1) Weak Emergence via Underivability- The emergent entity can be derived from its base properties, but only by iterating the systems’ microdynamic while taking the initial

starting conditions as inputs. That is to say, the emergent entity is derivable only by “simulation.”

This account of weak emergence is most associated with Bedau’s (2003a, 2003b, 1997) descriptions of artificial life (or Alife) simulations, in particular Conway’s famous “Game of Life” simulation. These Alife simulations consist of a grid in which each of the cells can be either alive (filled) or dead (empty). The grid follows simple update rules that govern what, given the current configuration of alive and dead cells, the cells will look like after a single tick of the game clock.⁹⁷ Though some configurations are simple enough that one can derive later configurations via computational shortcuts, other configurations can *only* be derived via simulation. That is, some later configurations can only be “discovered” through repeated application of the update rules on the Alife world as we find it. Bedau contends that in these underivable configurations have “weakly emergent” properties and underivability except by simulation is necessary for weak emergence.

For example, Bedau (2003a) discusses how the ‘R-pentomino’ configuration of cells (a configuration of five cells in a vaguely ‘R’ shape) starts off very simple yet evolves into an increasingly chaotic state (p. 20). At later states the configuration starts to generate cohesive combinations of cells that “move” across the grid (referred to as “gliders”) as they iterate living and dying patters. This raises a question, does the ‘R-pentomino’ configuration propagate indefinitely? The answer is ‘no,’ at the 1103rd iteration of the Alife world it settles into a stable state of cells that do not expand out or create further “traveling” configurations (p. 20). But,

⁹⁷ The four rules for Conway’s “Game of Life” are as follows:

For a space that is populated:

- (1) Each cell with one or no neighbors dies, as if by solitude.
- (2) Each cell with four or more neighbors dies, as if by overpopulation.
- (3) Each cell with two or three neighbors survives.

For a space that is empty or unpopulated

- (4) Each cell with three neighbors becomes populated.

Bedau argues, there is no way to *derive* this fact prior to simulating each step for that configuration of cells. Moreover, this is not a result of contingent human limitation. A Laplacian supercalculator presented with all of the initial information would *also* have to iterate each subsequent step of the system in order to know that the R-pentomino halts at the 1103rd iteration.

Here, then, is the objection: according to Bedau, *having a halting point at the 1103rd iteration* is an emergent property of the R-pentomino configuration. But, a filled cell in the Game of Life has the dispositional property of *forming a configuration whose development halts at the 1103rd generation when combined with other cells into an R-pentomino*. Since this is the property of a single cell, this is plausibly part of the micro-level, however, if it is included in the micro-level then the supposed emergent macro-property *is* derivable from the micro-level. Hence, the emergent property ‘collapses’ into the micro-level (Taylor 2015b, p. 737).

Of course, Taylor (2015b) notes that any observer faced with the R-pentomino configuration for the first time would be unlikely to know that the configuration halts when it does, however, such knowledge is beside the point. The question is whether, in fact, a filled cell *has* the dispositional property and thus it is irrelevant whether anyone *knows* if it has the property (p. 737). Having the property is all that is necessary for establishing the collapse objection. I now turn to a second example, this time of strong emergence:

(Ex. 2). *Strong Emergence as Novel Powers:* The emergent entity or property has a new causal capacity not found in its base properties or by a mere summation of the powers of these base properties—that is, the new property is non-structural.

A view like this has been predominately associated with O’Connor and Wong’s (2005) strong emergentist account as well as O’Connor’s (2000) earlier work on agent-causation. In both cases, the emergent property is understood as a non-structural natural property that is exemplified when systems attain an appropriate level of complex organization. The micro-properties that are so-

organized jointly cause the occurrence of the emergent property, which then non-structurally exhibits novel downward and linear causal powers (p. 111; O'Connor & Wong 2005, pp. 663-667). A property, *P*, is “structural” just in case the proper parts of entities with *P* themselves have properties that are not identical to *P* and jointly stand in a relation, and this state of affairs *just is* the entity having *P* (2000, p. 663). So, to say that the emergent property is non-structural is to say that its existence is over and above the composite of properties had by its base, that is, it is the claim that some *basic* or *fundamental* properties are had by *composite* individuals (p. 664).

This view is also subject to the collapse problem, perhaps even more obviously than in the previous weak emergence example. If the micro-properties already have, by O'Connor's lights, the disposition to *jointly cause the emergent property when organized in the appropriate way*, then it is reasonable that they also have the disposition *to exhibit the macro-level causal power via producing the emergent property when organized in the appropriate way*. But, if so, then the supposedly novel power of the emergent property *just is* one of the potentialities of the micro-properties. Hence, the strongly emergent property collapses to the micro-level.

In sum, emergent properties were thought to mark genuine levels in nature by establishing *real* distinctions between properties that trace nature's joints. The collapse problem suggests that emergence merely tracks arbitrary distinction between properties and offers no indication of real or significant differences.

The most natural way to resist this problem is to somehow restrict which properties can legitimately be included in the micro-level. If these ‘collapse inducing properties’ can be excluded for principled reasons, then the collapse problem is no longer applicable. Taylor (2015b) argues that all of the most plausible ways to do this fail by making the conditions for distinguishing between the micro-level and macro-level both too restrictive and too permissive.

Taylor (2015b) considers three distinctions one might use to try to exclude collapse properties from the micro level: (1) the *intrinsic/extrinsic* distinction, (2) the *dispositional/non-dispositional* distinction, and (3) *natural/non-natural* distinction (pp. 739-744). I briefly rehearse her responses. Regarding (1), the emergentist may want to restrict micro-properties to only include properties considered intrinsically. This is both too restrictive and too permissive. It is too restrictive because many micro-level scientific properties are *extrinsic*. For example, the property of *being a synaptic cleft* is an extrinsic relational property between two neurons that is fundamental to structures in neuroscience. Moreover, such a property plausibly operates as the micro-base in many emergent accounts of conscious minds. It is too permissive because, as many theorists hold, dispositional properties are often understood as intrinsic. So, even if we restrict the micro-level to include only intrinsic properties the properties that generate collapse in many cases would still be included.

Taylor (2015b) argues further that strategy (2), directly ruling out dispositional properties from the micro-level, also will not work. This strategy seems plausible, in part, because most of the collapse inducing properties were themselves dispositional. But, this tactic is, again, both too restrictive and too permissive. It is too restrictive because there are many natural micro-level properties that are considered dispositional and yet are not collapse inducing. For example, the property of *being a working neurotransmitter receptor*. This property has the requisite stimulus-response structure of dispositional property and yet it is a perfectly respectable neuroscience property without any danger of collapse. Eliminating dispositional properties would also be too permissive, since there are some properties that generate collapse yet are not dispositional. For example, Chalmers (1996; 2006) takes an emergentist view of consciousness, wherein all facts about the world are logically necessitated by fundamental physical facts *except* for facts about

consciousness. Instead, the fundamental physical facts in the base properties are related to emergent properties by psychophysical laws. But, Taylor (2015b) argues, this property of *being governed by the Psycho-Law*, if included at the micro-level, would mean that the micro-level *logically entails* the emergent properties (p. 738). Thus, Chalmers' proposed distinction collapses. Moreover, the property of *being governed by the Psycho-Law* is not a dispositional property, it has no clear stimulus-response type structure. Hence, even if we exclude dispositional properties from the micro-level, there will still be collapse inducing properties.

Finally, Taylor (2015b) considers (3), distinguishing between natural and non-natural properties. Natural properties are those properties that figure in natural laws or play a role in causation. They are thought to serve as the ground for other sorts of properties and are distinguished from non-natural amalgamations or mere Cambridge properties.⁹⁸ This attempt, Taylor (2015b) argues, *also* fails. It is too restrictive for two reasons. First, because the natural/non-natural distinction is usually understood as a spectrum, we would have to find some point on that spectrum to exclude properties from the micro-level (p. 743). This attempt runs the danger of being *ad hoc* if it merely targets properties known to generate collapse, but if it instead restricts it to *only* perfectly natural properties then it is far too restrictive since there are many plausible base properties that are not perfectly natural (e.g. certain structural properties). Second, the natural/non-natural distinction does not make good sense of examples like the Artificial Life case discussed above. In this case, none of the properties considered are natural full-stop, but if

⁹⁸ The term "Cambridge property" comes from discussions in the mid-20th century regarding the criterion of change. An initial proposal was that an object has change just in case it is true (false) that it has property P at time T₁ and false (true) that it has property P at time T₂. That is, change was thought to track the gaining or losing of properties. This, however, so-called "Cambridge criterion" was shown to be overly simplistic because an object can gain or lose properties without undergoing a change (i.e. Socrates gaining the property of *being taller than infant Plato* merely by virtue of Plato being born). These properties, which undermined the Cambridge account of change, became known as Cambridge properties, in part because of Geach's (1969) terminology that first highlighted the problem. For more a more detailed account of these properties and their features see Franciscotti (1999).

naturalness is context dependent than the lines of demarcation of even blurrier.

Moreover, this attempt is also too permissive. Taylor (2015b) argues that properties like those considered in the Chalmers example above (e.g. *being governed by a Psycho-Law*) are plausibly natural, but if so than at least some collapse inducing properties are still included (p. 743). Given these problems, Taylor (2015b) concludes that the only way to avoid the collapse problem is to shift discussion from properties in the world to the kind of *explanations* being offered in emergent accounts. Taylor argues that what is going wrong in examples of collapse is a violation of explanatory principles. This, then, points to a conception of emergences that is focused on a kind of explanatory failure, such as that elucidated in her other work (2015a).

As Taylor (2015a) argued, one reason to endorse her explanatory conception of emergence is that we already hold that there is a conceptual connection between emergence and explanatory failure (pp. 660-661). Here (2015b) she argues further that her conception provides a simple solution to the collapse objection (p. 748). Instead of restricting the collapse inducing properties from the micro-level base properties, she advocates a principle which excludes them from being part of the *explanandum* of an explanatory relation. She does this by arguing for the following minimal condition on explanation:

Let us say that two sentences are ‘factually equivalent’ iff there is no metaphysically possible world in which either sentence is true and the other false. So, for instances, ‘all ravens are black’ and ‘all non-black things are non-ravens’ are factually equivalent . . . I propose that no explanandum can be explained by a sentence to which it is factually equivalent. (2015b, p.749)⁹⁹

⁹⁹ She makes two further points here, first, that this is only a minimal constrain since it obviously takes a lot more for one sentence to explain another. Second, there is one important set of exceptions to this rule. In cases where the explanandum is metaphysically necessary, no attempt at explanation will meet the factual non-equivalence condition, however, one might think that certain metaphysical necessities still admit to certain kinds of explanations (e.g. mathematical explanations). But, even so, these exceptions should be small.

She calls this the “factual non-equivalence” criterion. This account and her explanatory construal of the concept of emergence (EM) provides all of the resources needed to prevent collapse problems from occurring. According to Taylor, the collapse inducing properties provide a powerful objection to traditional theories of emergence because there is no principled way to exclude them from the micro-level. In contrast, given the explanatory account of emergence provided by EM the only way a collapse objection could succeed is if the collapse inducing property provides an *explanation* which fills the explanatory gap. Call such an explanatory scheme a ‘collapse account.’ It is evident, however, that such collapse accounts violate the factual non-equivalence criterion.

For example, consider again Bedau’s (2003a) Artificial Life example. On the EM account the fact that the regularity ‘*whenever five living cells are combined into an R-pentomino, the resulting configuration halts at the 1103rd iteration*’ is without an adequate explanation is what grounds the claim that Alife simulations have weakly emergent properties. The previously considered collapse inducing property was *a living cell has the dispositional property of generating a configuration that halts at the 1103rd iteration when combined with four other living cells in a R-pentomino*. If this is *still* to be collapse inducing for an EM version of Alife emergence, then the collapse inducing property should figure in a collapse account that *explains* the purported explanatory gap. This explanation, however, takes the following form: “A living cell has the dispositional property of generating a configuration that halts at the 1103rd iteration when combined with four other living cells in a R-pentomino *and that explains why* living cells in a R-pentomino form a configuration that halts at the 1103rd iteration.” Clearly, this runs afoul of the factual non-equivalence criterion and as such this proposed property cannot induce

explanatory collapse, but since emergence *just is* a kind of explanatory failure on EM, neither can induce an emergent collapse.

In sum, Taylor (2015b) has further supported her account by both showing that all other metaphysically construed accounts are subject to the collapse problem and then showing that the collapse problem itself seems to point to the EM's account of emergence as the only viable alternative. I now consider two reasons to believe that Taylor's construal of the collapse objection fails.

Skiles (2016) criticizes Taylor's (2015b) solution to the collapse problem on two fronts. First, he argues that the explanatory principle regarding factual non-equivalence that she adverts to as a solution to the collapse problem is faces counter-examples. Second, he argues that there is still a successful version of the restriction strategy that takes the essential/non-essential distinction as a principled way to exclude collapse inducing properties from the micro-level. I briefly consider these proposals in reverse order.

Skiles (2016) argues that collapse can be avoided by appeal to the concept of something's "generic essence." In particular, the suggestion is that we can avoid the collapse of emergent property, *S*, by requiring that its micro-physical dependents, *P*, taken either individually or collectively do not have *S* as part of their generic essence (pp. 837-840). As Skiles (2016) puts it, "we demand the collection of features that constitutes a micro-level base of an emergent feature [*S*], be pure of [*S*]" (p. 840). Skiles (2016) uses Correia's (2006) specification of generic essence here in order to generalize the concept, and regimenting what it means for something to be "involved" in the generic essence of something else.¹⁰⁰ Skiles concludes that restricting the

¹⁰⁰ Some of the specifics are similar to Fine (2015) attempts to clarify the concept, which I discussed in chapter 3. These are different enough, however, that it is useful to briefly canvass the details of Correia (2006) account as used by Skiles. Correia's (2006) formulation is meant to capture both the generic essence of collections and individuals:

micro-level properties in this way is principled and avoids the difficulty of excluding too much or too little, which plagued other attempts (pp. 839-840). Taylor (2018a) responds to Skiles proposal by raising two potential objections to his response to the collapse objection. First, she argues that since Skiles admits that *tit* is part of the generic essence of the collection of features *being sodium* and *being chlorine* that when combined these create a compound that is water-soluble, then it seems there is no way to block the inclusion of many other features as part of the generic essence that collection (e.g. when combined these an ionic compound, when combined these create a salt that is white in colour, etc.). If so, however, then the generic essence becomes merely a list of true conditions about the collection of features, rather than an account of what it is to have those features (p. 392).

To truly capture the essence of *being sodium* and *being chlorine* we must instead focus on features by virtue of which those mere “true conditions” obtain, such as their respective atomic numbers. This, Taylor (2018a) argues, is a problem for Skiles because it shows we do not have a good reason to take “when combined creates a compound that is water-soluble” to be part of the generic essence of *being sodium* and *being chlorine*, as opposed to merely another true condition of those features (p. 392). As Taylor (2018a) concludes, “if water-solubility is not part of that generic essence, then properties involving water-solubility cannot be excluded from the micro-level base of water-solubility, and Skiles’ solution to the collapse problem is undermined” (p. 392). In like manner, Taylor (2018) argues that Skiles claim that “forming a compound that is water-soluble” is *not* part of the generic essence of *being sodium* but is part of the generic

$\exists F_1, F_2 \dots P$

Where this reads “what it is for an entity to have F_1 , some (perhaps different) thing to have $F_2 \dots$ is (at least in part) for the proposition that p to be true” (Skiles, 2016, p. 839) So, ‘ p ’ reports the generic essence of F_1 and F_2 as a collection. Importantly, this answers the question “what does it mean to *have* a given feature?” (generic essence), not “what *kind* of thing is the feature?” (objectual essence).

essence of the collection *being sodium* and *being chlorine* is tenuous at best (pp. 392-393). There seems to be no principled reasons for attributing the water-solubility to the collection but not the individual features.

Taylor's (2018a) objections to Skiles account raise good questions, but they seem to depend on an understandable yet unfortunate misreading of Skiles (2016). Contrary to Taylor's (2018a) claim that Skiles requires the collection of *Na* and *Cl* have the feature "forming a compound that is water-soluble" as part of their generic essence, Skiles (2016) states the opposite:

The distinction [between features taken individually or collectively] is important, as even if one resists that it is part of the generic essence of *Na* (i.e. *being sodium*) what occurs when it is bound with *Cl* (i.e., *being chloride*), and vice versa, one might nonetheless accept the following as a truth about the generic essence of the two features taken together:

$$\exists_{Na, Cl} \forall x \forall y (x \text{ is bound with } y \text{ to form NaCl} \rightarrow \exists z (z \text{ is water soluble}))$$

Presumably, though, the 'genuinely novelty' [sic] required for water solubility to be emergent—and thus, the proper way to formulate the ban against features that threaten it—would preclude water solubility from appearing in the very nature of being *Na* and being *Cl*, take **either** individually **or** collectively. (pp. 839-840, emphasis added)

Given this, Taylor's (2018a) objection to Skiles account fails to address the account as stated. Moreover, because of the severity of this misunderstanding even if we reconstruct Taylor's (2018a) objection so that it addresses Skiles' actual claim, it still does not succeed. After all, Skiles *would agree* that "being water-soluble" is merely a "true condition" of the collection *being sodium* and *being chlorine* precisely because "water-solubility" is the *target* emergent feature, and thus should be excluded from the generic essence of the micro-properties either taken individually *or* collectively.¹⁰¹

¹⁰¹ I hasten to add that Skiles' (2016) only explicitly uses examples that take the collapse properties to be attributed to *individual* micro-properties, even though he clearly implicates both individual *and* collective properties when he

There are still, however, reasons to reject Skiles (2016) response to the collapse problem. Baysan and Wilson (2017) argue that whatever merits it might have there remain two significant concerns. First, the notion of essence itself remains, for many, conceptually obscure and thus its specification is notoriously contentious. This is not a decisive reason to reject such a theory, however, since the collapse problem is a very *general* concern there is wisdom in seeking a more ecumenical approach (p. 77). Second, and more decisively, several going accounts of broadly scientific properties suggest that they are individuated by the laws they enter into. This entails that the essences of micro-level features of a purported emergent would fail to be “pure” of the emergent feature as Skiles requires, hence, collapse would remain a threat (p. 77).

What of Skiles’ (2016) negative argument *against* the “factual non-equivalence” (hereafter FN) criterion used by Taylor (2015b) to solve the collapse account? Skiles uses the famous example of a flagpole and its shadow, widely attributed to Bromberger (1966), to develop a counter-example to FN along the following lines:

$p_{\text{=def}}$ The flagpole’s height is 10 meters,

$p_c_{\text{=def}}$ The background conditions required to derive q from p ,

$q_{\text{=def}}$ The length of the flagpole’s shadow is 5 meters.

Skiles (2016) argues that the conjunction of $(p \wedge p_c)$ explains q . Moreover, if p_c is held fixed then p and q are interderivable. Thus, $(p \wedge p_c)$ and $(q \wedge p_c)$ are factually equivalent by Taylor’s definition since there is no metaphysically possible world in which either is true while the other is false. According to FN, a sentence cannot be factually equivalent to what it explains, yet, plausibly $(p \wedge p_c)$ partially explains $(q \wedge p_c)$. Hence, FN is false by counter-example (pp. 835-836).

generalizes the account. This choice of examples did Skiles no favors and, I suspect, contributed to Taylor’s (2018a) misreading.

Skiles (2016) notes that his objection has two key parts that might be questioned. First, he claims that FN is a requirement for partial explanation, since the same motivations that justified our acceptance of FN in the case of full explanation (e.g. avoiding circularity) also apply to partial explanation. Second, he proposes that any sentence that appears in a full explanation is a partial explanation (pp. 836-837). Hence, since $(p \wedge p_c)$ fully explains q , it should be a partial explanation of $(q \wedge p_c)$. If these two steps work, then it seems that Taylor (2015b) solution to the collapse problem also fails.

Taylor's (2018a) response to Skiles fares better than her objection to his positive account, but I argue that it should also be rejected. Taylor (2018a) gives two objections to Skiles counter-example. First, she claims that since p_c appears in both the *explanans* and the *explanandum* Skiles objection runs dangerously close to self-explanation. Given this, Taylor (2018a) argues, we should be suspicious of the claim that $(p \wedge p_c)$ at least partially explain $(q \wedge p_c)$ (p. 388). This response is too thin to go through. Skiles (2016) does have an argument for this claim, which involves iterating his earlier principle:

(PE) It is at least sufficient for a sentence (or sentences) to yield a partial explanation if conjoining it (or them) with other sentences yields a full explanation. (p. 836)

Given this, consider the following argument (Skiles 2016, p.837):

Given the following definitions:

p =def. The flagpole's height is 10 meters,

p_c =def The background conditions required to derive q from p ,

q =def. The length of the flagpole's shadow is 5 meters.

And the principle (PE)

(1) Let b = the conjunction of sentences $(S_1 \wedge S_2 \wedge S_3 \dots \text{etc})$, that together *fully explain* P_c

(2) If so, the conjunction $(p \wedge p_c \wedge b)$ together *fully explain* $q \wedge p_c$

Therefore:

(3) By (PE), $(p \wedge p_c)$ *partially explains* $(q \wedge p_c)$

Minimally, this shows that Skiles claim regarding the explanatory relation between $(p \wedge p_c)$ and $(q \wedge p_c)$ is supported by the sufficient condition for partial explanation, (PE), given above. As such, the self-explanation worry is less important than addressing (PE) directly, something that Taylor (2018a) goes on to do, which I discuss below.

We also might use this argument to avoid the self-explanation worry in the following manner. In the above schematic form, it seems that $(p \wedge p_c)$ fully explains ‘q’ and b fully explains p_c , which is why the conjunction $(p \wedge p_c \wedge b)$ together *fully explain* the conjunction $(q \wedge p_c)$. But, if so, then in the partial case what is being explained by $(p \wedge p_c)$ *just is* ‘q,’ however, since the statement under consideration is $(q \wedge p_c)$ it is a partial explanation since it only explains the ‘q’ part of $(q \wedge p_c)$. I am not sure Skiles would agree with this way of visualizing the explanatory relationships, but at the very least it would avoid any charge of self-explanation.

Leaving these speculations aside, Taylor (2018a) next argues against the Skiles’ underlying principle (PE) directly (pp. 388-389). Here she asks us to consider a fully physical explanation of a complex physical event, in this case Andy Murray winning the 2016 Wimbledon final. Such a full explanation will make reference to gravity and hence the following sentence will appear: “according to Newton’s law of universal gravitation, a particle attracts every other in the universe with a force directly proportional to the product of their masses and inversely proportional to the square of the distance between them.”¹⁰² Taylor (2018a) notes that *if* (PE) is true, then the above sentence must be a partial explanation of Murray’s win; however, intuitively

¹⁰² Of course, she is using Newtonian here for mere simplicity. A true full explanation would use a more sophisticated theory of gravity.

it is not even a partial explanation of the win (pp. 388-389). I take it the intuition tracks the thought the sentences about gravity, though necessary for a complete physical explanation of a given event, are too far removed to partially explain the same event.

I do not share this intuition. To be sure, the above sentence about gravity figures in explanations of almost all events, but this generality and “distance” from the target *explanans* does not mean it does contribute at all to partially explaining that event. I suspect what Taylor’s (2018a) intuition tracks is the salience of explanations in certain conversational contexts. It is true that in most everyday context when someone asks “why did Murray win?” appeal to gravity would not answer the question, even partially. This is not, however, because gravity is not a partial explanation of the *event* of Murray winning, objectively understood. Rather, it is because when asking that question what we usually mean is something like, “assuming all the normal background conditions, what additional factors explain Murray’s win?”

Taylor (2018a) recognizes the need to further support her intuition and appeals to contrastive explanations (p. 389). She claims that, “an explanation of Murray’s win would help us to understand why it occurred as opposed to not occurring” (p. 389). This need for a contrastive explanation, or as Taylor (2018a) puts it a favoring explanation, supports the intuition that sentences about gravity are not even a partial explanation for Murray’s win since gravity says nothing about why Murray won *rather than* lost (p. 389). Of course, as Taylor (2018) notes not all explanations are contrastive explanations, however, she claims that all is needed is *one* case to provide a counterexample to (PE) (p. 389, fn. 13).

I believe this claim is incorrect. Remember, Skiles posits:

(PE) If conjoining a sentence (or sentences) with other sentences yields a full explanation, then this is sufficient for it (or them) to yield a partial explanation.

It makes no claims regarding contrastive explanations. In order for the objection to go through it would have to be:

(PE*) If conjoining a sentence (or sentences) with other sentences yields a full explanation, then this is sufficient for it (or them) to yield a partial *contrastive explanation*.

I doubt, however, that Skiles would accept this principle. Adding the contrastive requirement to the antecedent of (PE) does not help either:

(PE**) If conjoining a sentence (or sentences) with other sentences yields a full contrastive explanation, then this is sufficient for it (or them) to yield a partial contrastive explanation.¹⁰³

This, even if a plausible principle, does not make good sense of the counter-example. For, it is unclear to me whether laws of gravity would figure in a “full contrastive explanation” of the event of Murray’s win. This is because, by my lights at least, “contrastive-explanation” serves to narrow the domain of explanation while “full explanations” serve to widen it. Thus, without a clear account of what a “*full* contrastive-explanation” amounts to, I am not sure how to evaluate Taylor’s counterexample.

These considerations show that Taylor’s (2018a) defense of her FN principle are insufficient. We should not, however, miss the broader point. The central contention between Skiles and Taylor, I suspect, revolves around how “metaphysical” explanations should be taken to be. Both Skiles and Taylor agree that some types of explanatory circles undermine explanatory power, but they disagree about how these vicious circles should be characterized. Skiles (2016) argument is about how Taylor’s (2015b) factual non-equivalence principle is insufficient for characterizing this constraint on explanation. Though Skiles does not frame it in this way, I believe his argument works because Taylor’s account ignores questions of *priority*.

¹⁰³ It seems that Taylor (2018a) has something like this in mind as a plausible requirement (p. 389).

That is, the debate between Taylor and Skiles mirrors debates concerning the relationship between dependence and modality. As I discussed in chapter 3, modal covariance may not yet say anything about what depends on what. In like manner, Taylor's (FN) principle is a modal principle, but it does not yet say *what explains what*. Put flatfootedly, the flagpole has explanatory priority over the shadow *even if* background conditions are held fixed such that statements about the shadow and statements about the flagpole were interderivable. Flagpoles explain shadows, not the other way around.¹⁰⁴

Of course, if Taylor's (2015b) attempt to solve the collapse problem fails for the reasons brought up by Skiles (2016) and if Skiles' own attempts at a solution are inadequate as Baysan and Wilson (2017) suggest, then we still need to address the collapse objection. Here I advert to the work of Baysan and Wilson (2017) who show, convincingly, that there are several well-motivated distinctions that would allow an emergentist to defend a metaphysically construed sense of emergence against various forms of the collapse objection, including Taylor's. Their argument is primarily aimed at defending various forms of *strong* emergence against collapse, but if a collapse defense of strong emergence is possible then weak emergence would not be hard to support. Baysan and Wilson (2017) argue that there are four distinctions that would allow for a principled defense from the collapse objection: (1) direct vs. indirect powers, (2) Lightweight vs. substantial dispositions, (3) Powers relativized to fundamental interactions vs. new powers, and (4) Strongly emergent objects. (p. 54).

The distinction between direct and indirect allows the emergentists to concede that, in some sense, the micro-physical level "inherents" the powers that are attributable to the macro-

¹⁰⁴ I suspect this style of critique, and Taylor's (2018a) recent responses, are motivated by a disagreement regarding the metaphysical robustness of explanation. Skiles seems committed to a more "realist" account whereby scientific explanation tells us something about the entities that explain, whereas Taylor seems more comfortable with a "anti-realist" or at least "pluralist" conception of explanation, which tends towards a more epistemic flavor.

level, but the emergentists then *deny* that these powers are had in the same way. The thought is that the micro-level base do not have the feature *directly* or in an *immediate* way, instead they only have it indirectly. Baysan and Wilson (2017) argue from analogy to temporally extended causal chains, where even though each link in the chain is sufficient for the next we can still distinguish between *direct* and *indirect* causes (p. 79). Just because there is a causal link between my birth and my lighting a fire in the fireplace does not mean that “lighting a fire” is not a novel event, nor does it mean that my birth has the power to light fires, except perhaps indirectly.

The second strategy, distinguishing between metaphysically lightweight and substantial dispositions, allows the emergentist to concede that the micro-level does have dispositions that bring about emergent features, but they then *deny* that these dispositions are metaphysically substantial. Instead, they are mere preconditions for the possession of the relevant emergent feature (2017, pp. 80-83). This avoids problems, noted by Taylor (2015b) that come from merely excluding *all* dispositions from the micro-level, instead it notes that some dispositions are different in *character*. Of course, the work now must go into determining whether or not a given disposition is lightweight or substantive, but *prima facie* the intuitive reaction that collapse inducing properties are somehow “cheap” or “missing the point” suggests that the burden for doing this lies with the proponent of collapse.¹⁰⁵

The third principled response involves a more stringently scientifically grounded distinction. Baysan and Wilson (2017) note that it is a scientific truism that powers are dependent

¹⁰⁵ Taylor (2015b) notes this same intuitive reaction to collapse properties in light of her own explanatory response. As she says, “we can see that this result [i.e. that collapse properties violate a basic explanatory principle regarding factual non-equivalence] accords with a common, intuitive response to collapse objections, which is something along the lines of ‘that’s cheating’ or ‘that’s clearly not what the author meant.’ The collapse objections generate dissatisfaction, and the connection between emergence and explanation reveals why this is the case . . . collapse objections simply bypass that explanatory work. A collapse objection is an attempt to establish ‘on the cheap’ that some property is non-emergent, and so it makes perfect sense to regard collapse objections as illegitimate.” (p. 751).

on, in some sense, fundamental forces or interactions.¹⁰⁶ For example, the power to bond to another electron is grounded in the electromagnetic interaction, as opposed to the strong nuclear interaction; or the power of a chair to uphold a human without them falling through is grounded in the gravitation and electromagnetic interactions jointly, etc. (pp. 83-84). The thought is that there are strong scientific reasons to think both that there are fundamental interactions/forces and the powers of ordinary objects, systems, and events are metaphysically dependent on such interactions (p. 84). If so, then we could define a metaphysical conception of emergences as whether the emergent property both depends on a physical base but also possess a power that is not dependent on any fundamental forces or interactions, that is to say, it would count as its own fundamental force or interaction.

This account clarifies what it would mean, metaphysically, for something to have a new *fundamental* power (as in strong emergence) and as such provides a principled reason for excluding certain higher-level features from the micro-base. The emergentist can concede that, relativized to fundamental interactions, the micro-level might inherit all the powers of any feature they necessitate, it could still be that composite features are associated with powers that are “new”, in not being grounded only in the set of physical fundamental interactions (pp. 84-89).

Finally, Baysan and Wilson (2017) argue that an emergentist could hold that strongly emergent features must be instantiated in a new object, different from the micro-level object’s base features (pp. 89-90). On this view, the collapse objection would plausibly be blocked because if no micro-level object is suited to be the bearer of the strongly emergent feature, where this feature distinctively involves a new power, then it is likewise plausible that the micro-level

¹⁰⁶ They also note that forces, roughly, involve pushes or pulls, whereas interactions include non-force goings-on, like exchanges or particle decays.

features had by the micro-level objects *also* do not have that power (p. 89). Effectively, the new object strategy flips the collapse objection. Instead of the possession of a disposition or property entailing a collapse from the macro-level to the micro, the possession of a property entails the expansion of salient objects and features.¹⁰⁷

We have considered two ways of unifying the concept of emergence, Wilson's (2015) "power-based" conception and Taylor's (2015a) "explanatory failure" account. Moreover, we examined at length the collapse problem, an objection that Taylor (2015b) raises against all current philosophical accounts of emergence especially those that are more metaphysically situated. Taylor (2015b) intends this problem to support an argument for her preferred conception of emergence as the only solution. I concluded that, in light of her debate with Skiles (Skiles, 2016; Taylor, 2018a), her solution to the collapse problem does not succeed. Further, even if Skiles' preferred solution *also* does not work, Baysan and Wilson (2017) have persuasively argued that there are several principled ways for an emergentist to respond.

Where, then, does this leave us? Even if the collapse problem can be addressed without endorsing the Taylor's (2015a; 2015b) account of emergence, there are other considerations that recommend her account, such as its conceptual unity and dialectic clarity. My preferred conception of emergence is a synthetic view that unites the dialectical clarity of Taylor's explanatory failure account, with the metaphysical import of Wilson's power-based account. I claim we should combine the two conceptual strategies for two reasons. First, given that emergence is used in many discursive contexts in both science and philosophy, there is strong reason for preferring a unitary conception of emergence as such. Moreover, because of the strong

¹⁰⁷ Baysan and Wilson (2017) also note that this strategy might work for weak emergences as well, at least under the "proper-subset" strategy (p. 96). In such a case, it may be that having a proper-subset of the micro-level powers might itself sometimes entail a new object to bear such a property.

conceptual connection between emergence and explanation, I agree with Taylor that the best *generic* characterization of emergence will focus on explanatory failure. Second, it is important to recognize that though emergence appears in many discursive contexts, these contexts share an important feature. Concepts of emergences are deployed within debates regarding reductive or non-reductive accounts. It is in *this* context that the concept of emergence was both developed and utilized, thus I agree with Wilson that having a concept with metaphysical *import* is crucial for capturing emergence as it is actually used.

Given the above considerations, I argue that emergence, as such, is related to strong emergence and weak emergence, as a genus is related to its species. Consider the following:

Generic Emergence (GE): A feature, F , is emergent from its microphysical base, B , where B is constituted by elements a, b, c, \dots etc. standing in relation R , just in case (i.) as a matter of nomiological necessity that F depends on B and (ii.) there is either no scientific or no metaphysical explanation for (i.) in terms of B .

Strong Emergence (Species I.): A feature, F , is strongly emergent if (GE) is true *because* F has one token power not identical with any token power of B on that occasion.

Weak Emergence (Species II.): A feature, F , is weakly emergent if (GE) is true *because* F has a non-empty proper subset of the token powers had by B , on that occasion.

These schema attempt to incorporate and relate the explanatory account of Taylor (2015a) with the power-based account of Wilson (2015). I have adjusted the explanatory account to remove the observer-relative notion that Taylor (2015a) emphasizes. Given that the concept of emergence is situated in debates about reductive and anti-reductive metaphysics, I concur with Wilson's (2015) claim that weak observer-relative emergence is not best understood as an account of emergence, but rather an *error theory* in support of reductive metaphysics (p. 396). Given that emergence was conceptually developed as an alternative to reduction, explicating a generic understanding of the concept in a manner that excludes such types of observer-relative

emergence is both useful and justified. Moreover, removing the observer-relative notion from the generic definition is consistent with my background assumptions of explanatory realism.

I have also adjusted the generic definition to include either scientific *or* metaphysical failures of explanation in an inclusive disjunct. This is meant to capture minimal cases of weak emergence that may be explainable *scientifically* at the micro-physical level, by a non-linear mathematical apparatus for example, yet are *ontologically* distinct, by Leibniz' law, and thus not *metaphysically* explained by the micro-physical level. As Wilson (2013) argues, certain kinds of non-linearity, such as when a higher-level entity exhibits fewer degrees of freedom than its base, are both "theoretically deducible" yet "ontologically irreducible" (pp. 227-229).¹⁰⁸¹⁰⁹

The emergent species are primarily applications of the strong and weak emergent schema proposed by Wilson (2015), I have taken them up with little change in part because I favor her more metaphysically robust construal of the concept. The only adjustment is in altering the first half of the schemas to reflect what makes them both belong to the emergent genus. Thus, the second half of these schemas (e.g. the "New Power Condition," and the "Subset of Powers Condition") becomes a *differentia* for the generic first half. This provides an answer to the question, in what sense are both of these *emergent*?" without losing the metaphysical specificity of an individual *species-level* account.¹¹⁰

¹⁰⁸ I am glossing over an important feature of Wilson's (2010) discussion, namely that there are actually three different ways that the degrees of freedom of higher-level special science entities might be characterized in contrast to the degrees of freedom of their micro-physical components: "'reductions,' 'restrictions,' and 'eliminations' in degrees of freedom" (p. 281). Strictly speaking, only "eliminations in degrees of freedom" support weakly emergent entities in Wilson's account.

¹⁰⁹ As a brief reminder, a "degree of freedom" is roughly "an independent parameter needed to characterize an entity as being in a state functionally relevant to its law-governed properties and behavior" (Wilson 2010, p. 281). The term is drawn from scientific discourse; however, it also has useful applications in metaphysics, metaphysics of nature, and philosophy of science. For an early explication of this concept, applied specifically to the relationship between supervenience and determinism, see Thalos (1999a; 1999b).

¹¹⁰ Given Wilson's (*forthcoming*; 2014) arguments *against* unnecessary unification and *for* concept pluralism in discussions of grounding and dependence, I suspect she might be against the modifications I have made. In response, I would point out that one salient difference between emergence and grounding is that emergence is

With this hybrid, genus/species account of emergence in hand, we can now turn to the primary task of applying this concept to the transitivity objection to my agency-first account of action. In the next section I articulate the objection more precisely, then I consider how the resources of emergence offer a way out of the objection. I then, in the final section, consider the *plausibility* of the sort of emergence needed to avoid the transitivity objection.

3. The Transitivity Objection and Emergent Agency

In chapter 3 I gave the following analysis of what makes an event count as an action:

Action as Essential Metaphysical Dependence (ver. 3): A mental event *e* is an action iff by virtue of the constitutive essence of interactive mental properties, *P*; *P* fully grounds all the properties brought about (or propositions made true) by virtue of the constitutive essence of *e*.

Part of the rationale for having *e*'s essence *fully grounded* in *P* rather than partially grounded was to ensure that this account explains action uniquely in virtue of its relation to agency. If such dependence was only partial and distributed, then my account would have failed to say anything interesting. But, a natural criticism is that my own use of the full grounding relation to describe how action-events relate to agents undermines this motivation. Grounding is, as I noted in chapter 3, traditionally understood to be a transitive relation. As such, if it is part of the essence of the properties that constitute our agency that they ground simple mental action-events and if, plausibly, our mental properties have *their* essence grounded in some lower-level features, then it follows that the simple mental action-event is also grounded in these lower-level features. Hence, if the relation of essential metaphysical dependence, construed as grounding, is meant to explain what makes an event count as an action, a critic can rightly ask why did I stop at the agential properties? Why not, instead, posit that what makes an event count as an action is that they are

explicitly used in scientific discourse. As such, the justification for attempting to provide a unified conception is, I believe, higher.

grounded in some micro-features that themselves ground agency? It seems, given transitivity, the ontological buck is passed down to the more fundamental supporting features and as such the agent, once again, disappears.

What is needed is a principled reason for taking the dependence of simple mental events on agential-properties *and exclusively* on agential properties to explain why an event counts as an action. I first consider two potential strategies for doing this that will not succeed before turning to my preferred response, which involves emergent mental properties.

An initially promising response involves the essential/existential distinction that I first investigated in chapter 3. Recall that, roughly, existential questions ask what is necessary for something to come into being, to begin or continue existing, whereas essential questions ask what is required for something to be the sort of thing it is, to be itself rather than some other kind of thing. These correspond to different types of dependence relations: existential dependence and essential dependence. The relationship between the conditions for something's existence and the conditions for something's essence are closely interconnected, but they can come apart. As I discussed, the dog breed, schnauzer, has many essential features (e.g. four leggedness, whiskers, stubborn disposition, etc.), which explain what makes a schnauzer be the sort of thing it is. None of these, however, explain the existence of schnauzers. Instead, the existence of schnauzers depends on certain facts outside of their essence (i.e. the political and social state of 15th century Germany, the need for rat-catching dogs in medieval Europe, the intentional actions of some breeders, etc.).

One might suggest that the transitivity objection can be avoided by claiming that while there is a transitive chain of existential dependence, essential dependence can in principle be separated from this. After all, I have already argued at length that essential dependence need not

collapse into existential dependence. If such a distinction can be maintained transitivity does not need to be invoked for the *essence* of actions, only their existence.

This response does not, however, address the objection. The criticism leveled at my account involves chains of *essential* dependence. Insofar as the agential properties of mind have essential features, it is appropriate to ask what makes these features be what they are, upon what do *they* essentially depend? If that question is answerable, then there is still a problem of transitivity, for whatever essentially grounds them would also essentially ground the simple mental events that are basic actions.

Another, perhaps more plausible response, is inspired by one of Baysan and Wilson's (2017) attempts to solve the collapse objection, is to distinguish between direct and indirect grounding (p. 67, fn 10; pp. 78-80). There are two ways to see this, first, suppose that necessity of origin is true and that this is an essential feature of entities. If so, then I essentially depend on the Big Bang. It seems that, even if this is true, it is not as salient for explaining upon what my thoughts essentially depend. If I answered the question, "what makes thoughts the sort of things they are" with "the Big Bang" this would, even if true, miss the point. Even if various features of me depend on distantly past events or fundamental laws they depend on these things only *indirectly* rather than directly.

The second way to make this point is inspired by Yates (2016) in his discussion of what he calls the "grounding objection" to physically acceptable strong emergence (pp. 834-837). Yates (2016) is arguing that we can make sense of strong emergence as an "in virtue" relation, wherein a given entity qualitatively realizes feature, *F*, wholly in virtue of its basic physical properties, *B*. The novel power, *P*, however, is solely exhibited in virtue of *F*, and it is necessary that *F*'s exhibit *P* (p. 835). In other words, *P* indicates that *F* is strongly emergent, in virtue of its

sole dependence on F . The critic, however, points out that since F is grounded in B , and P is grounded in F , it seems that P must *really* be grounded in B and thus P does not indicate emergence of any kind. The emergent power collapses.

Yates (2016) claims that the problem is that the critic is confusing the abstract logical relation of generic grounding (what Yates calls, “Grounding” with a capital ‘g’) with the *particular* grounding relations (or “little-g grounding”) involved between the B and F or F and P (p. 835). The key point is that the sense of “in virtue of” is different in each instance, in the first instance the “grounding” relation is one of “qualitative realization,” whereas in the second instances the “grounding” relation is one of “causal power conferral” (p.835). Both of these are instances of the “Grounding” relation, but they are distinct “g”grounding” relations. This distinction between relations offers a principled way for Yates (2016) to avoid the grounding version of the collapse objection.

Both Baysan and Wilson’s (2017) arguments regarding direct/indirect dependence and Yates’ (2016) argument about generic Grounding/particular grounding relations offer principled ways to mark breaks in chains of transitivity. Might they be useful for my purposes? Yes and no. I hold that both of these attempts, especially as they are situated in larger accounts of emergence, are on the right track. I have two concerns. First, Baysan and Wilson’s (2017) appeal to more or less direct dependence has vague boundaries, a fact they readily acknowledge (p. 80). Given that my project aims to answer what makes a simple mental event count as an action, it seems preferable to advert to a distinction with sharper boundaries if possible. Yates (2016) proposal is closer to what I would prefer, however, it is very specific to his account of strong emergence. Though I am, in general, sympathetic to strongly emergent accounts of mind (include broadly physicalist ones like Yates’), for a broad account of action like essential metaphysical

dependence, it would be methodologically preferable to not make it *depend* on a *particular* account of strong emergence. That being said, Yate's view is, roughly, on the right track. In what follows I will, using it as inspiration, apply it in conjunction with my hybrid account of emergence to solve the transitivity problem.

I claim that agential properties are plausible candidates for emergence, if anything is, though I will remain agnostic as to whether this emergence is weak or strong. Recall, that generic emergence is defined as:

Generic Emergence (GE): A feature, *F*, is emergent from its microphysical base, *B*, where *B* is constituted by elements *a, b, c, . . .* etc. standing in relation *R*, just in case (i.) as a matter of nomological necessity that *F* depends on *B* and (ii.) there is either no scientific or no metaphysical explanation for (i.) in terms of *B*.

So then, my claim is that the agential properties of mind as a matter of nomological necessity depend on some physical base properties (i.e. brain states, micro-physical states, etc.), but there is not a scientific or metaphysical explanation *in terms of those base physical properties* for the agential properties of mind. This statement of *GE* is already suggestive. In chapter 3 I noted that my preferred conception of metaphysical dependence construed grounding as *backing* metaphysical explanation, in the same way that causation backs causal explanations. Moreover, as I noted in the Introduction, I am presupposing explanatory realism, which claims that—all other things equal—explanations (or failures of explanation) tell us something *about what determines the explanandum*. Given these prior assumptions, the failure of explanation in *GE* with respect to agential properties *already* suggests a failure of grounding. In some sense, the agential properties are not wholly grounded in their base, because if they were they would be explainable *in terms of their base alone*.

We want to know, however, how this failure fits with my account of action. The species of *GE* are *Strong Emergence* and *Weak Emergence*:

Strong Emergence (Species I): A feature, F, is strongly emergent if (GE) is true *because* F has one token power not identical with any token power of B on that occasion.

Weak Emergence (Species II): A feature, F, is weakly emergent if (GE) is true *because* F has a non-empty proper subset of the token powers had by B, on that occasion.

I am following Wilson (2015) in taking “power” to be defined very broadly. Let’s examine how each species of emergence would apply to my account of action.

Species I. The agential properties are not explainable *in terms of* merely their physical base properties *because* the agential properties have the power to perform complex *actions*, which is not identical with any of the powers of the base properties. Moreover, agential properties have this power in virtue of how simple mental acts *are wholly essentially grounded* in these properties. Remember, my view is that simple mental acts (i.e. volitions) are basic actions, more complex actions have such basic actions as their causal source. So, then, the “power” had by agential properties *just is* the power to cause complex actions via volitions and this power is “had” in virtue of the agential properties fully grounding the essence of these volitions.¹¹¹

Given this account, I claim that the transitivity problem can be avoided in a manner similar to the collapse problem. That is, the sense in which the agential properties “depend” on their physical base is *different* than the sense in which a volition “depends” on the agential properties. It is this difference that, in part, marks out the agential properties *as* an emergent feature. My suggestion dovetails with Barnes’(2012) account of strong emergence.¹¹² She claims

¹¹¹ I am assuming a power can be “had” in this loose sense even if it is never actualized. So, in a world where an agent has volitions but is unable to perform any complex actions, the agential properties would still be emergent in virtue of the unique power that they have, which the agent (tragically) can never utilize.

¹¹² Strictly speaking Barnes (2012) is using considerations from debates in meta-ontology to characterize ontological emergence *in part* to say something about the usefulness of meta-ontology for questions in first order metaphysics. I largely ignore the nuances of this feature of her argument, instead focusing on her insights regarding strongly emergent properties.

that the thesis of strong emergence amounts to the claim that some things which are fundamental are also ontologically dependent. More generally, the metaphysical distinction between fundamental and derivative cuts orthogonally across the distinction of dependent and independent—if so, then if emergent entities exist, they are those entities that are fundamental yet ontologically dependent. The autonomy of the emergent entity is to be explained in terms of this fundamentality, especially one indication of fundamentality in nature is the existence of genuine, non-derivative, powers.

Earlier, in chapter 3, I proposed that a minimal condition of fundamentality is a kind of domain specific explanatory failure. There I suggested that agential mental properties were fundamental within the domain of the mental, drawing an analogy to baseline properties of consciousness. This means that these agential properties cannot be explained in terms of more fundamental *mental* properties that give rise to them, rather they can only be characterized by pointing to the other properties they support or give rise to, in this case simple mental events like volitions. This, in light of the connection between explanatory failure and emergence, both speaks to agential properties being *emergent* features and provides a solution to the transitivity objection. For, if there is a break in explanation then, given the connection between explanation and dependence, there is also a break in the transitivity of grounding.

This is not to deny that agential mental properties aren't, in some sense, dependent on their physical base for their existence. However, the part of those features that make them *fundamental* or *autonomous* would be, by definition, *not* explainable in terms of the base and so are *qua fundamental or autonomous feature* grounded in the physical base. Put differently, I deny that simple mental acts are grounded in the *dependent* part of the emergent property, but rather only the *autonomous* part. This provides a *principled* reason for denying the transitivity of

dependence with respect to simple mental events, they are still *wholly characterized* by their essential dependence on the agential properties and not the underlying base properties that support the essence of the agential properties themselves.

We might, naturally, want more details here. I wish to remain agnostic concerning the details of particular emergent relations, in part because I do not believe we know enough to be certain. But, here is a toy example that is plausible enough: Suppose that O'Connor and Wong's (2005) dynamic causal emergence account is, roughly, correct for characterizing how physical base properties produce a baseline emergent property. On this picture, then, agential properties would depend on their base properties in the sense of *causal* dependence. Once generated, however, the emergent properties then give rise to new features. For example, suppose that once the baseline emergent agential properties are generated they non-causally ground the essences of volitions, which confer the power of *acting* on agential minds. If so, then the fundamental and autonomous aspects of the emergent property, in virtue of which it is rightly called emergent, stand in a *different* relation (one of essential metaphysical dependence characterized by grounding) to those features that it generates, than it does to its own physical base. This explains how we can dismiss the transitivity objection in a principled way.

Species II. The agential properties are not explainable *in terms of* merely their physical base properties *because* the agential properties have the power to perform complex *actions*, which is a proper subset of the token powers had by their physical base on this occasion. Moreover, the agential properties have this power in virtue of how simple mental acts *are wholly essentially grounded* in these properties. It is natural to worry that weak emergence, unlike strong emergence, does not have the ontological resources to avoid the transitivity problem. I

argue, however, that even this minimal account of emergence can avoid the objection in a principled way.

Remember, the proper subset is meant to show how there can be genuine emergence without rejecting any standard physicalist desideratum. In this case, the weak emergentist denies that all instances of causal overdetermination are metaphysically problematic. As Wilson (2011) claims, the weak emergence necessary for non-reductive physicalism holds that certain special science entities (e.g. minds) are nothing over-and-above their physical base, that is they are physically realized (pp. 121-124). These special science entities, however, are also autonomous, that is, they are “*distinct from and distinctively causally efficacious as compared to—the lower-level entities upon which they depend*” (p. 121).

The proper-subset strategy holds that the higher-level entity is physically acceptable because it is nothing over-and-above its physical base, it is realized by its base and the novel powers it has *are also* possessed by a subset of the base. It is for this reason that the weakly emergent entity does not participate in problematic over determination. In one sense, both the physical base and the emergent entity are *causing* a given effect, however, the weak emergentist would claim that there is only *one power* on offer here. Thus, the sort of “overdetermination” (if that is even the right word) is not akin to problematic firing squad cases. As noted above, the autonomy, both ontological and causal, is provided by virtue of the distinctive *profile* that the special science entity “has,” which enters into special science explanations and causal interactions. If an entity possess a proper subset of its base’s powers, then it is ontologically distinct simply by Leibniz’s law (Wilson 2011, p. 129). It is distinctively causally efficacious *in comparison* to its base because it has a *distinctive causal profile*. The thought is that for the

emergent entity to be causally *distinctive* it does not require that it has a *new* power, it is enough that it have a distinctive set or collection of powers (p. 129).

One way to see this is to consider multiple realization. Following Wilson's (2011) example, suppose that a mental state, M, of "being tired" causes the effect, E, of desiring coffee and suppose M is realized by physical base P. Now, suppose it was also realized by *different* physical state P*, in relevantly similar conditions. Would M still cause E? Plausibly, yes, because the only powers that matter for E's occurrence are the *subset of powers* associated with M, all the other powers that differ between P and P* are irrelevant. That M's power profile (i.e. its collection or set of powers) is salient for E's occurrence is a good reason to think that M's power profile is distinctively efficacious *as compared* to P overall.

Moreover, these subsets of powers are what enter into special science laws and explanations, and it is plausible to think that if a particular subset of powers, had by an entity, is consistently referenced in the laws and explanations of a well-established science, this is sufficient reason for taking it to be tracking a distinctive causal contribution. This is meant to go hand-in-hand with the generic conception of "power" that Wilson is using, which we discussed above. Remember, the generic sense of "power" is just supposed to track the simple idea that what a particular can do depends on how it is—that is, what features, structures, or properties it has or is composed of. As Wilson (2011) explains, "it is in virtue of being massy, not magnetic, that a magnet falls to earth; in virtue of being magnetic, not massy, that a magnet attracts steel; and so on" (p. 126). So, then, the emergent property in the weak case *has* a distinctive power, not because it is new, but because it *has* a distinct collection of powers, and in light of the special science entity being this way (rather than the physical base being that way) it can make the causal contributions that it does. This larger point about the possession of powers—even if this

is merely the possession of a collection, set, or plurality—underwrites the claim Baysan and Wilson (2017) make regarding the potential of a “new object” strategy for avoiding the collapse objection, for it might be that sometimes the existence of a new object is required to bear the distinctive subset of the powers of the base entities (p. 96, see also pp. 92-93, and p. 93, fn 23).

How then does all of this fit with my account of action such that it provides a principled reason for rejecting the transitivity objection? On this view, the agential properties are only weakly emergent, this means that they are physically realized by some base entity (e.g. brain states), but they possess a distinctive *subset* of the powers of this physical base. In this case, it is the subset of powers associated with causing complex actions, but this subset is “had by” the agential properties *in virtue* of their fully grounding simple mental events—volitions. This is meant to track the plausible claim that the entity that is associated with the subset of powers both is not identical with its base (by Leibniz’s law) and is causally distinctive and that the conjunction of these facts suggests that the bearer of this proper subset is a *new*, though admittedly only weakly emergent, entity.

This limited sort of novelty is enough to avoid the collapse objection, but I argue it avoids the transitivity objection as well. For, in this case, while the physical base broadly grounds the existence of the agential properties, it does not ground, *qua physical base as such*, the essence of the agential properties. For, these essential features of the agential properties, the kind of features that in turn fully ground simple mental acts like volitions, are only grounded *in a proper subset* of the physical base, but this subset is not, as argued above, the same as the physical base. Rather, this subset *alone* just is the essential ground of the agential properties and their subsequent powers. This means that there is, in some sense, a complete scientific explanation for the distinctive powers of the emergent entity, after all, it is a subset of the

physical base. But, there is still a failure of *metaphysical explanation* in terms of the physical base, because we must posit an ontologically and causally distinctive entity to “track” this proper subset and the contributions it makes to causal outcomes. This is consistent with my account of the genus “emergence,” or *GE*, which requires that there is a failure of *either* scientific *or* metaphysical explanation. My claim, then, is that this break in metaphysical explanation provides a principled reason for thinking there is a break in essential dependence as well, given the close conceptual connection between grounding and explanation and my background assumption of explanatory realism.

Put differently, there is nothing scientifically unexpected regarding the actual powers possessed essentially by the agential properties in virtue of their grounding volitions, which is to say that there is nothing scientifically surprising about their existence. What *makes them essentially* the kind of thing they are, however, requires reference to the emergent level entity in order to “fix” the essential features, but if this “fixing” presupposes the emergent entity then it is not really *essentially* explainable in terms of its physical base alone. That is, the agential properties are contextually fundamental at the level of mental properties and thus weakly emergent.

Interestingly, this weak emergence interpretation of my account of action has some resemblance to Ginet’s (1990) response to claims that physical brain events are the non-actional cause of volitions, which then causally explain actions (pp. 9-10). Against this claim, he notes that any brain processes that, plausibly, underwrites our actional mental processes are *already* actions themselves, thus they cannot explain what makes something count as an action by the mere fact that they causally underwrite volitions (pp. 10-11). Because of this, Ginet (1990) famously claims that *even if* volitions are identical with certain neurological causal processes it is

not *in virtue* of these causal processes that they count as actions instead it is something intrinsic to the simple mental event, which has no further causal structure (pp. 12-13). In like manner, the weak emergence view articulated above supposes *even if* agential properties have powers *identical* to a proper subset of the powers of its base, this cannot explain *what it is* for these properties to be what they are. Instead, we must look at what they essentially do *at the level of mental properties* to characterize them and it is at this level that *in virtue* of their essential grounding of simple mental events that these agential properties possess the powers they do.

4. Conclusion—A Question of Plausibility?

The above considerations show that, though initially challenging, the transitivity objection can be addressed in a principled way by applying the concept of emergence. Moreover, my account can remain agnostic regarding whether strong emergence or weak emergence best characterizes the relationship of the mental to the physical.¹¹³ In either case, the emergence of agential properties answers the transitivity objection by providing principled reasons for holding that a mental property is either all-things-considered fundamental, as in strong emergence, or contextually fundamental, as in weak emergence. Furthermore, this fundamentality goes hand-in-hand with a kind of explanatory failure, either scientific or metaphysical, which provides strong reasons to doubt the continuity of *essential* dependence relations. Even if the above account is correct, however, there is a final lingering question: “is such a view plausible?” Emergence, of all sorts, has been subject to much doubt concerning its scientific respectability. Thus, even if this is a possible solution, one might worry it is not a plausible solution for any theory that wishes to remain scientifically respectable.

¹¹³ As an aside, I note that I am sympathetic to a strongly emergent account of the mental, especially the agential features of mind, which plausibly possesses new and *fundamental* powers.

Answering this concern *fully* would be beyond the scope of this project, however, I partly support the plausibility of my appeal to emergence by canvassing the various ways that emergence is returning to scientific discussions. The upshot of these examples is simply that emergence, of both strong and weak verities, is a *live* option at nearly all levels of scientific study. Moreover, the seriousness of these options is reflected in the lively debates between scientific reductionist and scientific emergentists.

I take Gillet's (2016) monograph on scientific emergence as my point of departure, and most of the examples I use are either referenced by or related to research programs that he mentions. I should also note before engaging with specific examples that I am deliberately eliding some details about how both the scientists and philosophers of science I mention are using the concept. Undoubtedly, sometimes the conceptualization of emergence used in one research program will contradict that used in another domain. Moreover, Gillet's (2016) own notion does not neatly fit with the conception of emergence I propose above. Again, the aim of this conclusion is merely to establish that emergent interpretations of scientific data and theories is a viable *scientific* interpretation and not merely an invention of philosophers. Merely indicating the existence of the debate, without clarifying the details of every account is sufficient for this aim.

The explanation for why emergence has returned to scientific discourse is disputed. Biophysicist Harold Morowitz (2002), famous for his work on the application of thermodynamics to living systems and the emergence of life, argues that scientific emergence has its roots in the early development of statistical mechanics, which was then accelerated by the rise of both information and complexity studies—aided by computers—in the mid-20th century (pp. 9-14). Gillett (2016) points to theoretical physicist Philip Anderson's seminal 1972 article

on the tension between symmetry breaking in physics and certain reductionist paradigms of explanation as a “first salvo” in the present set of emergentist debates (p. 3, fn 6).¹¹⁴

Regardless of the genesis of contemporary scientific reduction, there is no doubt that it has come to play an important role in contemporary theorizing. So much so, that Robert Laughlin (2005), a prominent condensed matter physicist, has boldly proclaimed:

Much as I dislike the idea of ages, I think a good case can be made that science has now moved from an Age of Reductionism to an Age of Emergence, a time when the search for ultimate causes of things shifts from the behavior of parts to the behavior of the collective. (p. 208)

This is, I think, perhaps a little too hasty. As Gillet (2016) rightly observes, the question of whether reduction or emergence is the right scientific research paradigm is still an open question; in part because more empirical evidence is needed or because even where evidence is suggestive, there has not been enough attention paid to plausible rival explanations (pp. 322-323). However, given statements such as McLaughlin’s (1992) conclusion that though emergentism is rationally coherent its defenders do not have a “scintilla of evidence” for their claims, the mere fact that there exists a *lively* debate in various sciences is a far better empirical situation for supporters of emergentism than the first half of the 20th century provided (p. 90-91).¹¹⁵ It is with an eye towards this overall *contested* yet scientifically lively situation that I provide the following, suggestive examples.¹¹⁶

¹¹⁴ Philip Anderson (1972). “More is Different: Broken Symmetry and the nature of the Hierarchical Structure of Science.” *Science*, 177, pp. 393.

¹¹⁵ Indeed, so lively it has appeared in congressional testimony. The funding debates regarding the so-called “supercollider” in the late-80’s and mid-90’s involved, among other things, disputes among the scientific experts regarding if the project was worth the large amount of governmental money. Philip Anderson famously testified *against* the collider. This testimony is sometimes characterized as exemplifying the deep scientific conflict between high-energy physicists and condensed matter physicists in America, however, some recent studies have also pointed to Anderson’s aforementioned emergentist commitments, whereas the research paradigm of the proponents of the collider tend to have a more reductionistic paradigm. See, Martin (2015) for more historical details about these methodological, social, and scientific conflicts. For an interesting account from the *other* side of these funding debates, see Weinberg (1987) republished in Bedau and Humphreys (2008).

¹¹⁶ Though I mention several different scientific examples, many of them are taken from physics. This is, in part, because contemporary debates regarding scientific emergence *started* with condensed physics. I think the more

(i.) *Physics, Complexity Studies, and Computer Science*

I have already mentioned Anderson's (1972) initial work on emergence and "symmetry breaking," this area continues to be a fruitful line of inquiry for emergentists.¹¹⁷ Laughlin and Pines' (2000) general treatment of spontaneous symmetry breaking in the context of condensed matter physics is an accessible discussion of why, as they claim, such principles lead to "protectorates" within physical theories that are governed by higher-level organizing principles and nothing else (pp. 261-262). Batterman (2011) attempts to flesh-out the implications of such a theory—building also on Batterman (2002) and (2000)—particularly with an eye towards how details of the mathematical apparatus can be used to explain the *existence* of such emergent "protectorates" from the more fundamental symmetries without denying their emergence. For a recent, very technical, account of the limits of these implications, however, consider Landsman's (2013) account of spontaneous symmetry breaking in quantum systems.

The concept of symmetry breaking historically developed in condensed matter physics and many of the sources mentioned above tend to use, among other things, examples from condensed matter physics. Laughlin (2005) is, again, a good source for information about this, in particular he focuses on the phenomenon of superconductivity. By his lights, the properties of superconductors are novel, precisely because they exhibit features that would not exist if we merely look at the underlying quantum mechanical laws. For a philosophical interpretation of this Gillet (2016) and Auyang (1998) are both good sources.

important feature, however, is that philosophers, including philosophers of science, tend to privilege physics. This, for better or worse, shaped the available accessible sources I used.

¹¹⁷ Symmetries are, roughly and generally, defined as invariances across some set of transformations. There are a host of physical phenomenon that exhibit such symmetries (e.g. the rotational symmetry of space-time). Symmetry breaking, then, is when a phenomenon depends (or is describable) only by "breaking" a broader symmetry group to a lower-level symmetry, one of its sub-groups. The type of symmetry breaking that is most encouraging for emergentists is, so-called, "spontaneous symmetry breaking" where there are solutions to equations pertaining to a given symmetry that are not invariant with respect to that symmetry, even though there is no outside asymmetric input, hence the spontaneity.

In the above examples, part of the theoretical apparatus used to make sense of these higher-level phenomenon is the technique of “renormalization”—see Batterman (2011, pp. 1042-1044) for a good explanation of the procedure—and the associated loss of “degrees of freedom.”¹¹⁸ We have already seen the concept of degrees of freedom used to great effect by Wilson (2011), however, Thalos’ (2006) account is useful in applying it more directly to some of the physical theories we have been discussing, including thermodynamic behavior at “critical states,” as also discussed in Batterman’s (2002;2000) accounts. One important difference in Thalos’ (2006) discussion is that whereas Wilson and Batterman focus on “eliminations” of degrees of freedom as a sign of emergence, she discusses the “conferring” of degrees of freedom as a more viable (and more properly non-reductive) alternative.¹¹⁹ Taking the theme of thermodynamics a step further, but tying it together with questions of explanatory fictions and issues in quantum mechanics, Shech (2018) has provided a recent study that, though technical, offers interesting conclusions about the viability of certain claims of scientific *strong emergence* with respect to quantum phenomenon.

Turning more explicitly to quantum mechanics, Humphreys’ (1997) account of the relation between entangled quantum states and emergence is a good representative account from a philosophy of science perspective. It has been critically assessed and expanded upon by Kronz and Tiehen (2002), who offer several alternative interpretations of what emergence, in light of quantum mechanics, must look like. More broadly, Effective Field Theory (EFT)—a contemporary development of Quantum Field Theory—has been assessed as a locus of

¹¹⁸ If one wants a deeper dive into questions of the exact nature of renormalization explanations as an explanation of universalities see Franklin (2018)

¹¹⁹ Though Thalos does not frame it this way, I wonder if this difference between accounts of DOFs might track a meaningful distinction between conceptions of weak vs conceptions of strong emergence.

emergence, Cao and Schweber (1993) being a representative example.¹²⁰ Tying these quantum considerations back to the issues arising in condensed matter physics, Franklin and Knox (2018) have argued for a novel case of scientific emergence in the case of phonons (a “quasiparticle” associated with the collective excitation of arrangements of atoms and molecules in condensed matter) and underlying quantum mechanical laws. It is also worth mentioning Wilson’s (2013) intriguing proposal that the discovery of the weak nuclear interaction can be characterized as the discovery of “fundamental novelty” in nature and thus provides a model of scientifically respectable strong emergence (p. 211).¹²¹

Finally, taking a broader view, Coleman (2018) and Blundell (2016)—both physicists—have attempted to provide holistic accounts of how emergence functions methodologically in physics. Blundell (2016) takes emergent features to primarily be a function of the limited subjective awareness of finite minds and thus it serves as a kind of “storytelling” for making sense of reality, however, he seems to reject the conclusion that this makes emergence “merely” epistemic, instead favoring a hybrid approach.¹²² Coleman (2018), in contrast, suggests a reconciliatory account that takes both reduction and emergence to be an ineliminable part of physics, but where reduction is *primarily* methodological and merely provides the mathematical techniques and tools for understanding genuine emergent reality.

¹²⁰ For a critical assessment of their strongly antireductive claims, see Castellani (2002). She argues that EFTs do not prove as much as Cao and Schweber think, though she remains agnostic about a summery verdict in the pro-reduction direction either.

¹²¹ I should also, for the sake of completeness, mention Crull’s (2013) interesting proposal that quantum decoherence actually undermines the whole dialectic of *both* reduction and emergence.

¹²² I confess that I find his reconciliatory conclusion philosophically puzzling, but I mention it because it is coming from a physicist and thus serves to substantiate my claim that emergence is considered a live option within physics. Blundell (2016) seems to claim that because human minds are embedded in the natural world we should take these epistemic limitations to be, in some sense, an indication of what is real. As he describes, “emergent laws and properties are independent, novel structures that function effectively because they are well adapted for human thought processes” (p. 2) and again, “But isn’t this just weak emergence? Are not scientists simply struggling with their imperfect models and wrestling with questions of epistemology, rather than addressing reality head on? I reject such a clear-cut distinction. Emergent properties are members of the set of elements of reality, and as such merit ontic status” (p. 11).

Many of the considerations above are linked with nonlinearity and complexity studies. Of course, nonlinearity has already been mentioned in the context of Wilson's (2013) work, but it is worth noting again here the development of such ideas in Alife simulations and other synthetic life projects. Bedau's (2013; 2003a; 2003b; 1997) accounts are the most commonly cited examples, but Varenne et al. (2015) takes these ideas a step further and applies them to simulated social or group lifeforms. Perhaps the best resource for discovering a host of these nonlinear examples in sciences ranging from condensed matter physics all the way up to biology is Scott (2007).

(ii.). Biology and Bio-Systems

I have already mentioned Morowitz (2002) the biophysist who concluded that his work on the origins of life and thermodynamics provides evidence for several kinds of emergent properties. Most of the recent arguments regarding emergent properties in the life sciences come of systems biology, which takes the functioning of wholistic biological systems as the fundamental subject of inquiry. O'Malley and Dupré (2005) provide a good overview of the relationship between systems biology and emergent properties.¹²³ Gillet (2016) has pointed to several specific examples of this as well, including Boogerd et al. (2005) who argue that biochemical networks in eukaryotic cells provide evidence of emergence and Richardson and Stephan (2007), who apply it to the regulation of diauxic growth in *Escherichia coli*. Mosso, Bich, and Moreno (2013) have, more generally, worked to provide an account of inter-level causal relationships in biology, describing this this as a kind of physically respectable emergent causal power.

¹²³ For an interesting discussion about background challenges for systems biology and what metaphysics of nature best supports it see Bertolaso and Ratti (2018).

Wilson and Holldobler (1988) have a classic study regarding eusocial insects (mostly ants), which is notable because Wilson was initially a scientific reductionist yet he seemed to accept some sort of emergent phenomenon in these cases. Wilson (1998) would later renounce this acceptance of emergence and return to a more full-throated reductionism, but the reasons for both of his changes of mind are enlightening. The evidence from eusocial insects has been picked up by other scientists and philosophers of science, with Mitchell (2009; 2012) providing a good representative example this time focused on bees. Gillet (2016) has a useful overview of the whole development of these insect studies.¹²⁴

(iii.). Psychology and Cognitive Science

Psychology and Cognitive Science are both scientific domains that are close to the most contentious debates about emergence in philosophy. As such, it is useful to remind ourselves that regardless of the trenchant debates in philosophy of mind the concept of emergence appears in viable scientific debates. Of course, philosophers have long pointed to the failure of a reductive behaviorist research program as evidence that the mind might not be well modeled by a reductive paradigm (e.g. Burge, 2010a; 2010b), but it is true that the language of emergence itself is not often explicitly used in psychology. Antonietti (2010) has offered a general analysis of why this might be the case, along with a helpful survey of how the concept might be clarified within psychology. Walmsley (2010) has, in like manner, offered an analysis of how emergence relates to the so-called “dynamical approach” to cognitive science. Finally, Runyan’s (2014) monography on neural cause of human action provides an interesting scientific and philosophical account of emergence and its relation to agency.

¹²⁴ I also briefly note that discussions about the relationship between evolutionary development *itself* and emergence, which were popular in the late 1800s and early 1900s (e.g. the biological and philosophical work of de Chardin), might be a viable contemporary research program (see Salmon 2009, and Bedau’s 2018 APA-Pacific talk) but since I know of no current research programs along those lines I leave this as a footnote.

(iv.) *Final Considerations*

This *very* limited list of scientific examples (and philosophy of science interpretations) is both rough and incomplete, however, it provides a helpful antidote to a general emergence skepticism.¹²⁵ Of course, this does not *establish* the applicability of emergence to reality, like Gillet (2016) I want to caution restraint about making such claims with any certainty. It is enough for my argument to show that emergence is a *viable* scientific interpretation. Moreover, I note in passing that *if* we think, for independent reasons, that my account of action is the most plausible and it requires emergence, then we would have some defeasible justification for accepting the emergence of mental properties.

¹²⁵ Gillet (2016, p. 351, fn 18) mentions several examples in addition to those already brought-up, including: Prigogine's (1968; 1997) work on thermodynamical systems; Garfinkel (1987) and Goldbeter's (1997) work on chemical signaling in slime molds; Freeman (2000a; 2000b) on neural populations; Couzin and Krause (2003) regarding flocking behavior in animals, and Camazine et al. (2001) with a survey of many examples of biological self-organization, etc. For full citations see Gillet (2016).

CONCLUSION

Where does this leave us? I have argued that what makes a simple mental event count as an action is as follows:

Action as Essential Metaphysical Dependence (ver. 3): A mental event e counts as an action iff by virtue of the constitutive essence of interactive mental properties, P ; P fully grounds all the properties brought about (or propositions made true) by virtue of the constitutive essence of e .

This picture was motivated by two considerations: (1) causal conceptions of action tend to be reductive and for this reason lose sight of the agent in trying to describe actions. This often leads to counter-examples, as in causal deviance cases, and where causalists avoid these counter-examples they often must smuggle broader non-causal features of agency back into their account. (2) Most non-causal theories of action are opaque regarding exactly *what is* the relationship between agent and action. Non-causalists often focus on phenomenal features as either an explanation for what counts as an action or evidence that actions are uncaused. Relying on such arguments puts causalists and non-causalists into a dialectical stalemate that not only is unproductive, but also contributes to the unpopularity of non-causal views among philosophers of action.

Unlike these standard views, my account tied simple mental actions directly to agents by appealing to the non-causal yet extrinsic relation of essential metaphysical dependence. Moreover, such an account meets many of the desideratum of weak non-causalism (i.e. it is non-reductive, compatible with actions being caused, intrinsically active, etc.) and addresses crucial objections from causal theories (i.e. it involves an extrinsic relation, preserves agential control, etc.). Finally, I addressed the plausibility of my account, especially in light of the transitivity objection, by examining the concept of emergence and how it applies to action. I concluded that even a minimal account of emergence would prevent the transitivity objection from succeeding.

Whether or not mental properties are *strongly* emergent is a matter of much debate, however, the claim that they are at least *weakly* emergent is more widely held. Moreover, given the myriad of live debates in both metaphysics and science about emergent properties in nature, I take the sort of emergence that allows my account to avoid the transitivity objection is, in fact, plausible.

The previous chapters do not provide a *conclusive* argument for my essential metaphysical dependence account; however, I do think they articulate a plausible alternative to current theories of action, which resolves the dialectical stalemate between causalists and non-causalists. I conclude my dissertation by gesturing towards two avenues of future research that my account of “essential metaphysical dependence” invites: (1) *free action*, and (2) *animal action*.¹²⁶

1. Free Action and Essential Metaphysical Dependence

First, and most obviously, the topic of *free action*. In my dissertation I have carefully abstracted away from libertarian and compatibilist concerns, even though many traditional non-causalists are motivated by their libertarian sensibilities. A complete theory of action must deal with free action and essential metaphysical dependence is no exception. I highlight, roughly, what shape such developments might take and note the outstanding challenges for crafting a libertarian version of essential metaphysical dependence—my preferred view.

¹²⁶ In this conclusion I only examine two future avenues of research, however, I briefly note that *divine action* is another subject to which my account could be fruitfully applied. Traditional puzzles in philosophy of religion about how God could act in the natural world could benefit from a view of action that does not place (in the first instance) an act within the causal nexus but rather in how an event depends or is grounded in a mind—potentially a God-like mind. Such an account has precedent, for example, Aquinas explanation of miracles (as I understand it) shares some features. According to Aquinas, a miracle does not imply that God breaks into or intervenes upon causal chains within the world. Instead, because all entities depend on God for their being (essence), God “acts” by fully realizing a feature of their being (essence) that was previously obscured. Though clearly different in many ways, such an account is similar to mine insofar as it looks to a relationship of essential dependence as the basis for (some of) God’s actions.

First, an account of action in terms of metaphysical dependence suggests that dependence relations, in general, may be useful for understanding traditional problems of free will. Several philosophers of action and metaphysicians have started using the notion of dependence to talk about traditional problems in action theory. For example, Tognazzini's (2015) argument regarding the luck objection (the objection that *indeterminism* undermines freedom by making agency "random"), which frames the strongest form of the luck objection *in terms* of what grounds actions. Indeed, though Ginet (2014, 2007) never uses the term "ground" or "dependence" explicitly, his engagement with the luck objection suggests that he understands it as a question of dependence. Such uses of the dependence idiom have also appeared in discussions of freedom and foreknowledge, such as Swenson's (2016) treatment of divine foreknowing and freedom.¹²⁷ Swenson's (2016) work is doubly useful since, though it is aimed at the foreknowledge and freedom debate, in it he also provides an account of "ability" in terms of dependence. This works suggest that the time is ripe for philosophers to explore how dependence and traditional problems of free will might relate.

After establishing the salience of dependence for questions of freedom, the next step would be to give a positive account of free agency. I believe that such an account could be presented as a contrast case. I would contrast my agency-first, grounding based, account of action with Ginet's (1990) non-causal conception of free agency, on the one hand, and O'Connor's (2001) agent-causal account, on the other. Since my account of action shares elements with both Ginet and O'Connor's, it is natural to try and frame an account of free agency in light of their contributions. It is difficult to see, ahead of time, how such an account might unfold, however, several challenges are obvious from the beginning. An account of free action in

¹²⁷ For an extended look at the metaphysics of foreknowledge which uses the concept of dependence but attempts to more precisely state what this dependence might look like, see Anderson and Watson (2010).

terms of metaphysical dependence, if it is going to be a libertarian view as I would prefer, would need to distinguish itself from compatibilism. As my view of action is currently framed, the weak non-causalism combined with the necessity of the grounding relation might legitimately raise questions regarding, in what sense, the agent is free. Of course, something like a grounding interpretation might suggest source incompatibilist views, such as those recently defended by Pereboom (2014). I see two potential ways forward: First, following Tognazzini's (2015) work on the luck objection, I could attempt to reformulate the deterministic threat in terms of *grounding* and thus make incompatibilism via a failure of dependence more plausible. Second, following O'Connor's (2005) work on the distinctive *limitations* of human freedom, I could talk about how questions of dependence and sourcehood might be related to more traditional alternative possibility incompatibilist views. Regardless of the success of such arguments, it seems that there is room to explore what freedom as *uncaused* essential metaphysical dependence might look like.

2. Animal Action and Essential Metaphysical Dependence

Philosophers of action are, as a rule, focused on human action. Such accounts appeal to relatively complex mental entities like intentions with propositional content, or sophisticated conscious experiences like the "actish phenomenal quality." All of these views, as plausible as they might be, tellingly avoid the question of non-human animal action. Some moral philosophers have taken up this question in the context of trying to understand the nature of rationality/agency's relationship to morality—most notably Korsgaard (2018)—however, most have not dwelt extensively on this topic from a purely metaphysical perspective.

This blind spot in philosophy of mind, philosophy of action, and metaphysics is beginning to be filled. In the past 15 years or so there are several works, especially in psychology

and philosophy of mind, that are taking the status of animal agency seriously (e.g. Lazareva, 2012; Burge, 2009; Allen, 2006; Hurley 2003). There is, however, no extensive treatment of animal agency *from the perspective* of philosophy of action. My account, in contrast to traditional theories of action, does not focus on sophisticated mental entities or phenomenal qualities, but rather an extrinsic relation of dependence between minds and simple mental events. As such, it leaves the exact *manner* or *characterization* of the properties involved open to specification. Insofar as my theory of action does not, from the outset, overintellectualize what makes an event an action there is space to use my view as a starting place for discussing animal action.

Moreover, O'Connor (2005) argues that we can clarify the concept of free agency by framing human action against the upper limit-case of divine action (or what divine action must be, if it exists). This method is admirable, but it should be extended by considering the *lower* limit-case—non-human animal action. Clarifying what it means for animals to act, aside from its intrinsic interest, allows us to better understand the place of human action in the world. Furthermore, once agency is seen as primary it gives us a helpful frame for understanding *how* agency develops in the natural world.

Finally, investigating non-human animal agency from the philosophy of action perspective allows us to raise new questions that have been largely ignored. For example, few have considered possibility and the scope of non-human animal freedom. Though, as with free human action, one cannot prejudge how these arguments will unfold, I suspect that animal freedom is both more common and more extensive than is often thought by philosophers. Supposing this is correct, if one is a libertarian and if one wants to maintain the plausible claim that most animals are not morally responsible, then this recommends a novel inverse of

Fischer's (2007) "semi-compatibilist" view. Freedom—of a certain sort—is necessary but not *sufficient* for moral responsibility.

Developing my account in these directions situates human action within a broader continuum of agency that *includes* non-human animals. This both clarifies what is *unique* about human action (and free action) while, at the same time, revealing how it fits into the larger natural world. Overall, these two considerations—free action and animal action—show that my account of action has room to grow into a full and fruitful non-causal theory of agency.

LIST OF REFERENCES

- Adams, Frederick and Alfred Mele. (1992). "The Intention/Volition Debate." In *Canadian Journal of Philosophy*. Vol. 22, No 3. 323-337.
- Aguilar, Jesús H. (2012) "Basic Causal Deviance, Action Repertoires, and Reliability." *Philosophical Issues*. Vol. 22. 1-19.
- Allen, Colin. (2006). "Transitive Inference in Animals: Reasoning or Conditioned Associations?" In *Rational Animals?* Edited by Susan Hurley and Matthew Nudds. New York: Oxford University Press.
- Alvarez, Maria. (2018). "Reasons for Action, Acting for Reasons, and Rationality." *Synthese*. Vol. 195. 3293-3310.
- Anderson, David J. and Joshua L. Watson. (2010). "The Mystery of Foreknowledge." *Philo*. Vol. 13, No. 2. 136-150.
- Anderson, Philip. (1972). "More is Different: Broken Symmetry and the nature of the Hierarchical Structure of Science." *Science*, 177, pp. 393.
- Anscombe, G. E. M. (1971) *Causality and Determination: An Inaugural Lecture*. Cambridge: Cambridge University Press.
- Antonietti, Alessandro. (2010). "Emerging Mental Phenomena: Implications for Psychological Explanations." In *Emergence in Science and Philosophy*. Edited by Antonella Corradini and Timothy O'Connor. New York: Routledge. 266-288.
- Audi, Robert. (2012). "A Clarification and Defense of the Notion of Grounding." In *Metaphysical Grounding: Understanding the Structure of Reality*. Edited by Fabrice Correia and Benjamin Schnieder. New York: Cambridge University Press.
- _____. (1993). *Action, Intention, and Reason*. Ithaca, NY: Cornell University Press.
- Auyang, Sunny. (1998). *Foundations of Complex-Systems Theories in Economics, Evolutionary Biology, and Statistical Physics*. Cambridge: Cambridge University Press.
- Barnes, Elizabeth. (2012). "Emergence and Fundamentality." *Mind*. Vol. 121, No. 484. 873-901.
- Baron, Sam and James Norton. (2019). "Metaphysical Explanation: The Kitcher Picture." *Erkenntnis*. Published Online (16 January 2019). <https://doi.org/10.1007/s10670-018-00101-2>.
- Batterman, Robert W. (2011). "Emergence, Singularities, and Symmetry Breaking." *Foundations of Physics*. Vol. 41. 1031-1050.
- _____. (2002). *The Devil in the Details: Asymptotic Reasoning in Explanation, Reduction, and Emergence*. New York: Oxford University Press.

- _____. (2000). "Multiple Realizability and Universality." *The British Journal for the Philosophy of Science*. Vol. 51. 115-145.
- Baysan, Umut and Jessica Wilson (2017) "Must Strong Emergence Collapse?" *Philosophica*. Vol. 91. 49-104.
- Bedau, Mark A. (2013). "Weak Emergence Drives the Science, Epistemology, and Metaphysics of Synthetic Biology." *Biological Theory*. Vol. 8, No. 4. 334-345.
- _____. (2010). "Weak Emergence and Context-Sensitive Reduction." In *Emergence in Science and Philosophy*. Edited by Antonella Corradini and Timothy O'Connor. New York: Routledge. 6-46.
- _____. (2008). "Is Weak Emergence Just in the Mind?" *Minds and Machines*. Vol. 18, No. 4. 443-459.
- _____. (2003a). "Downward Causation and Autonomy in Weak Emergence." *Principia*. Vol. 6, No. 1. 5-50.
- _____. (2003b). "Artificial Life: Organization, Adaptation, and Complexity from the Bottom Up." *Trends in Cognitive Sciences*. Vol 7, No 11. 505-513.
- _____. (1997) "Weak Emergence." *Philosophical Perspectives*. Vol. 11. 375-399.
- Bennett, Karen. (2017). *Making Things Up*. New York: Oxford University Press.
- Bernstein, Sara. (2016). "Grounding is Not Causation." *Philosophical Perspectives*. Vol. 30. 21-38.
- Bertolaso, Marta and Emanuele Ratti. (2018). "Conceptual Challenges in the Theoretical Foundations of Systems Biology." In *Systems Biology*. Edited by Mariano Bizzarri. New York: Springer. 1-13.
- Bird, Alexander. (2007). "The Regress of Pure Powers?" *The Philosophical Quarterly*. Vol. 57, No 229, 513-534.
- Bittner, Rudiger. (2001). *Doing Things for Reasons*. New York: Oxford University Press
- Blanchard, Thomas. (2016). "Physics and Causation." *Philosophy Compass*. Vol. 11. 256-266.
- Block, Ned. (1995) "On a Confusion about a Function of Consciousness." *Brain and Behavioral Sciences*. Vol. 18, No. 2. 227-247.
- Blundell, Stephen J. (2016). "Emergence, Causation, and Storytelling: Condensed Matter Physics and the Limitations of the Human Mind." Accessed June 2018. *ArXiv*, 1604.06845.

- Boogerd, F.C., F.J. Bruggeman, R. C. Richardson, A. Stephan, and H. V. Westerhoff. (2005). "Emergence and its Place in Nature: A Case Study of Biochemical Networks." *Synthese*. Vol. 145. 131-164.
- Brewer, Talbot. (2009). *The Retrieval of Ethics*. New York: Oxford University Press.
- Broad, C. D. (1952). *Ethics and the History of Philosophy*. London: Routledge & Kegan Paul.
- Boyd, Richard. (1980). "Materialism Without Reductionism: What Physicalism Does Not Entail." In *Readings in the Philosophy of Psychology, Vol 1*. Edited by Ned Block. Cambridge, Mass.: Harvard University Press. 1-67
- Burge, Tyler. (2010a) *Origins of Objectivity*. New York: Oxford University Press.
- _____. (2009). "Primitive Agency and Natural Norms." *Philosophy and Phenomenological Research*. Vol. 79, No. 2. 251-278.
- _____. (2010b). "Modest Dualism." In *The Waning of Materialism: New Essays*. Edited by Robert C. Koons and George Bealer. Oxford University Press.
- Campbell, C. A. (1951). "Is 'Freewill' a Pseudo-Problem?" *Mind*. Vol. 60, No. 240. 441-465.
- Cao, Tian Yu and Silvan S. Schweber. (1993). "The Conceptual Foundations and the Philosophical Aspects of Renormalization Theory." *Synthese*. Vol. 97, No. 1. 33-108.
- Castellani, Elena. (2002). "Reductionism, Emergence, and Effective Field Theories." *Studies in History and Philosophy of Modern Physics*. Vol. 33. 251-267.
- Chalmers, David. (2006). "Strong and Weak Emergence." In *The Re-emergence of Emergence: The Emergentist Hypothesis from Science to Religion*. Edited by Philip Clayton and Paul Davies. Oxford: Oxford University Press. 244-255.
- _____. (1996). *The Conscious Mind*. New York: Oxford University Press.
- Clarke, Randolph. (2010). "Because She Wanted To." *The Journal of Ethics*. Vol. 14. No. 1. 27-35.
- Clark, Michael J. and David Liggins. (2012). "Recent Work on Grounding." *Analysis Reviews*. Vol. 72, No. 4. 812-823.
- Coleman, Piers. (2018). "Emergence and Reductionism: An Awkward Baconian Alliance." Forthcoming in *Handbook of Emergence*. (2019). Edited by Sophie Gibb, Robin Hendry, and Tom Lancaster. Accessed February 2018. *ArXiv*, 1702.06884v1.
- Corradini, Antonella and Timothy O'Connor. (2010). "Part I: Introduction." In *Emergence in*

- Science and Philosophy*. Edited by Antonella Corradini and Timothy O'Connor. New York: Routledge. 3-7.
- Correia, Fabrice. (2013). "Metaphysical Grounds and Essence." In *Varieties of Dependence*. Edited by Manuel Hoeltje, Benjamin Schnieder, and Alex Steinberg. Munich: Philosophia Verlag. 271-291.
- _____. (2010). "Grounding and Truth Functions." *Logique et Analyse*. Vol. 53. 251-279.
- _____. (2008). "Ontological Dependence." *Philosophy Compass*. Vol. 3, No. 5. 1013-1032.
- Correia, Fabrice and Alexander Skiles. (2017). "Grounding, Essence, and Identity." *Philosophy and Phenomenological Research*. Published Online (20 October 2017). Accessed January 2019. doi: 10.1111/phpr.12468.
- Correia, Fabrice and Benjamin Schnieder. (2012). "Grounding: an opinionated introduction." In *Metaphysical Grounding: Understanding the Structure of Reality*. Edited by Fabrice Correia and Benjamin Schnieder. New York: Cambridge University Press.
- Colyvan, Mark. (2001). *The Indispensability of Mathematics*. New York: Oxford University Press.
- _____. (1998). "Can the Eleatic Principle be Justified?" *Canadian Journal of Philosophy*. Vol. 28, No 3. 313-335.
- Correia, Fabrice. (2008). "Ontological Dependence," *Philosophical Compass*. Vol. 3, 1013-1032.
- _____. (2006). "Generic Essence, Objectual Essence, and Modality." *Noûs*. Vol. 40, No 4. 753-767.
- Crull, Elise. (2013). "Exploring Philosophical Implications of Quantum Decoherence." *Philosophy Compass*. Vol. 8, No. 9. 875-885.
- Dasgupta, Shamik. (2017). "Constitutive Explanation." *Philosophical Issues*. Vol 27. 74-97.
- _____. (2014). "On the Plurality of Grounds." *Philosophers' Imprint*. Vol. 14, No. 20. 1-27.
- Davidson, Donald. (1963) "Actions, Reasons, and Causes." *The Journal of Philosophy*. Vol. 60, No. 23, APA- Eastern Division, 685-700.
- _____. (1973) "Freedom to Act." In *Essays on Freedom of Action*. Edited by Ted Honderich. London: Routledge.

- _____. (1970). "Mental Events." In *Essays on Actions and Events*. Edited by L. Foster and J. W. Swanson. Oxford: Clarendon Press. 207-224.
- _____. (2001). *Essays on Actions and Events*. New York: Oxford University Press.
- Davis, Lawrence H. (1979). *Theory of Action*. Englewood Cliffs, New Jersey: Prentice-Hall.
- Dennett, Daniel. (1991). "Real Patterns." *Journal of Philosophy*. Vol. 88. 27-51.
- _____. (1984) *Elbow Room: The Varieties of Free Will Worth Wanting*. Cambridge, Mass.: MIT Press.
- Dickenson, Jason. (2007). "Reasons, Causes, and Contrasts." *Pacific Philosophical Quarterly*. Vol. 88. 1-23.
- D'Oro, Giuseppina and Constantine Sandi. (2013). "From Anti-Causalism to Causalism and Back: A History of the Reasons/Causes Debate". In *Reasons and Causes: Causalism and Anti-causalism in the Philosophy of Action*. Edited by Giuseppina D'Oro and Constantine Sandis. New York: Palgrave Macmillan.
- Dray, W. H. (1957). *Laws and Explanation in History*. Oxford: Clarendon Press.
- Fine, Kit. (2015). "Unified Foundations for Essence and Ground." *Journal of the American Philosophical Association*. Vol. 1, No. 2. 297-311.
- _____. (2012a). "Guide to ground," in *Metaphysical Grounding: Understanding the Structure of Reality*. Edited by Fabrice Correia and Benjamin Schnieder. New York: Cambridge University Press.
- _____. (2012b). "Pure Logic of Ground." *The Review of Symbolic Logic*. Vol. 5, No. 1. 1-25.
- _____. (1995a). "Ontological Dependence." *Proceedings of the Aristotelian Society*. Vol. 95. 269-290.
- _____. (1995b). "The Logic of Essence." *Journal of Philosophical Logic*. Vol. 24, No. 3. 241-273.
- _____. (1994a). "Essence and Modality: The Second Philosophical Perspectives Lecture," *Philosophical Perspectives*. Vol. 8, 1-16.
- _____. (1994b). "Senses of Essence." In *Modality, Morality, and Belief: Essays in Honor of Ruth Barcan Marcus*. Edited by Walter Sinnott-Armstrong, Diana Raffman, and Nicholas Asher. New York: Cambridge University Press.
- Fischer, John Martin. (2007). "Semi-Compatibilism." In *Four Views on Free Will*. New York: Wiley-Blackwell.

- _____. (2006). *My Way: Essays on Moral Responsibility*. New York: Oxford University Press.
- _____. (1994). *The Metaphysics of Free Will: An Essay on Control*. Walden, Mass. : Wiley-Blackwell.
- Fischer, John Martin and Mark Ravizza. (1999) *Responsibility and Control: A Theory of Moral Responsibility*. New York: Cambridge University Press.
- Fodor, Jerry A. (1974) "Special Sciences: Or, The Disunity of Science as a Working Hypothesis." *Synthese*. Vol. 28, No. 2. 97-115.
- Francescotti, Robert. "Mere Cambridge Properties." *American Philosophical Quarterly*. Vol. 36, No. 4. 295-308.
- Franklin, Alexander. (2018). "On the Renormalization Group Explanations of Universality." *Philosophy of Science*. Vol. 85, No. 2. 225-248.
- Franklin, Alexander and Eleanor Knox. (2018). "Emergence Without Limits: The Case of Phonons." *Studies in History and Philosophy of Science: Part B: Studies in History of Modern Physics*. Vol. 64. 68-78.
- Geach, P. T. (1969). *God and the Soul*. London: Routledge.
- Gillett, Carl. (2016). *Reduction and Emergence in Science and Philosophy*. New York: Cambridge University Press.
- Ginet, Carl. (2016). "Reason Explanation: Further Defense of a Non-causal Account." *Journal of Ethics*. Vol. 20. 219-228.
- _____. (2014). "Can an Indeterministic Cause Leave a Choice up to the Agent?" in *Libertarian Free Will: Contemporary Debates*. Edited by David Palmer. New York: Oxford University Press.
- _____. (2008). "In Defense of a Non-Causal Account of Reasons Explanations." *The Journal of Ethics*. Vol. 12. 229-237.
- _____. (2007). "An Action Can be Both Uncaused and Up to the Agent." In *Intentionality, Deliberation and Autonomy: The Action-Theoretic Basis of Practical Philosophy*. Edited by Christoph Lumer and Sandro Nannini. New York: Routledge. 243-255.
- _____. (2002). "Reasons Explanations of Action: Causalist Versus Noncausalist Accounts," in *The Oxford Handbook on Free Will, 1st Edition*. Edited by Robert Kane. New York: Oxford University Press.
- _____. (2000). "Book Review: The Works of Agency: On Human Action, Will, and

- Freedom by Hugh McCann” *The Philosophical Review*. Vol. 109, No 4. 632-635.
- _____. (1990). *On Action*. New York: Cambridge University Press
- Goetz, Stewart. (2008). *Freedom, Teleology, and Evil*. London: Continuum.
- _____. (1997). “Libertarian Choice.” *Faith and Philosophy*. Vol. 12. No. 2. 195-211.
- _____. (1988). “A Noncausal Theory of Agency.” *Philosophy and Phenomenological Research*. Vol. 49, No. 2. 303-316.
- Hall, Ned. (2004). “Two Concepts of Causation.” In *Causation and Counterfactuals*. Edited by John Collins, Ned Hall, and Laurie Paul. Cambridge, Mass.: MIT Press. 225-276.
- Hempel, C. G. (1942). “The Function of General Laws in History”, *Journal of Philosophy*. Vol. 39.
- Hitchcock, Christopher. (2007). “Three Concepts of Causation.” *Philosophy Compass*. Vol 2. No. 3. 508-516.
- Hornsby, Jennifer. (1980). *Actions*. London: Routledge.
- Horgan, Terence. (1989). “Mental Quausation.” *Philosophical Perspectives*. Vol. 3. 47-76.
- Hume, David. (1773) *An Enquiry Concerning Human Understanding*. Edited by Tom L. Beauchamp. Oxford: Oxford University Press, 1999.
- Humphreys, Paul. (1997). “How Properties Emerge.” *Philosophy of Science*. Vol. 64. 1-17.
- Hurley, Susan. (2003). “Animal Action in the Space of Reasons.” *Mind and Language*. Vol. 18, No. 3. 231-256.
- Hutto, Daniel. (1999). “A Cause for Concern: Reasons, Causes, and Explanations.” *Philosophy and Phenomenological Research*. Vol. 59, No. 2. 381-401.
- Kane, Robert. (1996). *The Significance of Free Will*. New York: Oxford University Press.
- Kim, Jaegwon. (2006). “Emergence: Core Ideas and Issues.” *Synthese*. Vol. 151, No 3. 547-559.
- Korsgaard, Christine. (2018) *Fellow Creatures: Our Obligations to the Other Animals*. New York: Oxford University Press.
- Koslicki, Kathrin. (2016). “Where Grounding and Causation Part Ways: Comments on Schaffer.” *Philosophical Studies*. Vol. 173, No. 1. 101-112.
- Kripke, Saul. (1972). *Naming and Necessity*. Cambridge, Mass.: Harvard University Press.

- Kronz, Frederick M. and Justin T. Tiehen. (2002). "Emergence and Quantum Mechanics." *Philosophy of Science*. Vol. 69, No. 2. 324-347.
- Lange, Marc. (2016). *Because Without Cause: Non-Causal Explanations in Science and Mathematics*. New York: Oxford University Press.
- Laughlin, Robert. (2005). *A Different Universe: Reinventing Physics from the Bottom Down*. New York: Basic Books.
- Laughlin, Robert and David Pines. (2000). "The Theory of Everything." *Proceedings of the National Academy of Sciences of the United States of America*. Vol. 97, No. 1. 28-31. Reprinted in *Emergence: Contemporary Readings in Science and Philosophy*. Edited by Mark A Bedau and Paul Humphreys. (MIT Press, 2008). 259-268.
- Lazareva, Olga, Toru Shimizu and Edward Wasserman (eds). (2012). *How Animals See the World: Comparative Behavior, Biology, and Evolution of Vision*. New York: Oxford University Press.
- Lewis, David (1983). "Extrinsic Properties", *Philosophical Studies*. Vol 44, 197–200.
- Lipton, Peter. (1991). "Contrastive Explanation and Causal Triangulation." *Philosophy of Science*. Vol. 58., No. 4. 687-697.
- Lloyd, Elisabeth. (1995) "Feminism As Method: What Scientists Get That Philosophers Don't." *Philosophical Topics*. Vol. 23, No. 2. 189-220.
- Lyons, Timothy D. (2003) "Explaining the Success of a Scientific Theory." *Philosophy of Science*. Vol. 70, No. 5. 891–901.
- Martin, Joseph D. (2015) "Fundamental Disputations: The Philosophical Debates that Governed American Physics, 1939-1993." *Historical Studies in the Natural Sciences*. Vol. 45, No. 5. 703-757.
- Maurin, Anna-Sofia. (2018). "Grounding and Metaphysical Explanation: It's Complicated." *Philosophical Studies*. Published Online (17 March 2018). Accessed January 2019. <https://doi.org/10.1007/s11098-018-1080-0>.
- McCann, Hugh J. (2012). "Making Decisions," *Philosophical Issues*. Vol. 22, 246-263.
- _____. (1998). *The Works of Agency: On Human Action, Will, and Freedom*. Ithaca: Cornell University Press.
- McCann, Hugh J. and Jonathan L. Kvanvig. (1991). "The Occasionalist Proselytizer: A Modified Catechism." *Philosophical Perspectives*. Vol. 5. 587-615.
- McDonnell, Neil. (2015). "The Deviance in Deviant Causal Chains." *Thought: A Journal of*

Philosophy. Vol. 4. 162-170.

McKenna, Michael. (2013). "Reasons-Responsiveness, Agents, and Mechanisms." In *Oxford Studies in Agency and Responsibility, Vol. 1*. Edited by David Shoemaker. 151-183.

McLaughlin, Brian P. (1992). "The Rise and Fall of British Emergentism." In *Emergence or Reduction?: Prospects for Nonreductive Physicalism*. Edited by Has Flohr and Jaegwon Kim. New York: Walter de Gruyter. 49-93.

Melden, A. I. (1961). *Free Action*. London: Routledge.

Mele, Alfred. (2017). *Aspects of Agency: Decisions, Abilities, Explanations, and Free Will*. New York: Oxford University Press.

_____. (1992). *Springs of Action*. New York: Oxford University Press.

Mitchell, Sandra. (2012). "Emergence: Logical, Functional and Dynamical." *Synthese*. Vol. 185. 171-186.

_____. (2009). *Unsimple Truths: Science, Complexity and Policy*. Chicago, IL: University of Chicago Press.

Morowitz, Harold. (2002). *The Emergence of Everything: How the World Became Complex*. New York: Oxford University Press.

Mossio, Matteo, Leonardo Bich, and Alvaro Moreno. (2013). "Emergence, Closure, and Inter-Level Causation in Biological Systems." *Erkenntnis*. Vol. 78, No. 2. 153-178.

Nolan, Daniel. (2014). "Hyperintensional Metaphysics." *Philosophical Studies*. Vol. 171. 149-160.

Nagel, Ernest. (1952). "Wholes, Sums, and Organic Unities." *Philosophical Studies*. Vol. 3, No. 2. 17-32.

O'Brien, Lilian. (2012). "Deviance and Causalism." *Pacific Philosophy Quarterly*. Vol. 93. 175-196.

Ó Conaill, Donnchadh and Tuomas Tahko. (2018) "New Frontiers in Ground, Essence, and Modality: Introduction." *Synthese*. Published Online (17 December 2018). Accessed January 2019. <https://doi.org/10.1007/s11229-018-02067-8>.

O'Connor, Timothy. (2000). *Persons and Causes: The Metaphysics of Free Will*. New York: Oxford University Press.

_____. (2005). "Freedom with a Human Face." *Midwest Studies in Philosophy*.

Vol. 29. 207-227.

- _____. (1994). "Emergent Properties." *American Philosophical Quarterly*. Vol. 31. 91-104.
- O'Connor, Timothy and Hong Yu Wong. (2005). "The Metaphysics of Emergence." *Noûs*. Vol. 39, No. 4. 658-678.
- Oliver, Mary. (2006). *Thirst*. Boston: Beacon Press.
- O'Malley, M. and Dupré, J. (2005). "Fundamental Issues in Systems Biology." *Bioessays*. Vol. 27, No. 12. 1270-12706.
- Palmer, David. (2016). "Goetz on the Noncausal Libertarian View of Free Will." *Thought: A Journal of Philosophy*. Vol. 5. 99-107.
- Papineau, David. (2002). *Thinking About Consciousness*. New York: Oxford University Press.
- Plantinga, Alvin. (1981). "Is Belief in God Properly Basic?" *Nous*. Vol.15, No 1.
- Pereboom, Derk. (2014). *Free Will, Agency, and Meaning in Life*. New York: Oxford University Press.
- Putnam, Hillary. (1994). *Words and Life*. Cambridge, Mass.: Harvard University Press.
- _____. (1973). "Reductionism and the Nature of Psychology." *Cognition*. Vol.2, No 1. 131-46. Reprinted in *Words and Life* (1994). 428-40.
- _____. (1967). "The Mental Life of Some Machines." In *Intentionality, Minds and Perception*. Edited by Hector-Neri Castañeda. Detroit, MI: Wayne State University Press.177-200. Reprinted in *Mind, Language and Reality* (1975). 408-28.
- _____. (1963). "Brains and Behavior." In *Analytical Philosophy, Second Series*. Edited by R. J. Butler. Oxford: Basil Blackwell. 1-19. Reprinted in *Mind, Language and Reality* (1975). 325-41.
- Raven, Michael J. (2015). "Ground." *Philosophy Compass*. Vol. 10, No. 5. 322-333.
- Rawls, John. (1951). "Outline of a Decision Procedure for Ethics." *Philosophical Review*. Vol. 60, No 2. 177-197.
- _____. (1971) *A Theory of Justice*. Cambridge, Mass.: Harvard University Press.
- _____. (2001) *Justice as Fairness: A Restatement*. Edited by Erin Kelly. Cambridge, Mass.: Harvard University Press.

- Richardson, Robert C, and Achim Stephan. (2007). "Mechanisms and Mechanical Explanation in Systems Biology." In *Systems Biology: Philosophical Foundations*. Edited by F. Boogerd, F. Bruggeman, S. Hofmeyr, and H. Westerhoff. Amsterdam: Elsevier. 123-144.
- Rodríguez-Pereyra, Gonzalo. (2005). "Why Truthmakers?" In *Truthmakers: The Contemporary Debate*. Edited by Helen Beebe and Julian Dodd. New York: Oxford University Press.
- Rosen, Gideon. (2010). "Metaphysical Dependence: Grounding and Reduction," in *Modality: Metaphysics, Logic, and Epistemology*. Edited by Bob Hale and Aviv Hoffmann. New York: Oxford University Press.
- Ruben, David-Hillel. (1987). "Explaining Contrastive Facts." *Analysis*. Vol. 47, No 1. 35-37.
- Runyan, Jason D. (2014). *Human Agency and Neural Causes*. London: Palgrave-Macmillan.
- Salmon, James F. (2009). "Emergence in Evolution." *Foundations of Chemistry*. Vol. 11, No. 1. 21-32.
- Salmon, Wesley. (1984). *Scientific Explanation and the Causal Structure of the World*. New York: Princeton University Press.
- Schaffer, Jonathan. (2016a). "Grounding in the image of causation." *Philosophical Studies*. Vol. 173, 49-100.
- _____. (2016b). "Ground Rules: Lessons from Wilson." In *Scientific Composition and Metaphysical Ground*. Edited by Kenneth Aizawa and Carl Gillett. New York: Palgrave-Macmillan. 143-169.
- _____. (2012) "Grounding, Transitivity, and Contrastivity." *Metaphysical Grounding: Understanding the Structure of Reality*. New York: Cambridge University Press. 122-138.
- _____. (2009). "On what grounds what," in *Metametaphysics: New Essays on the Foundations of Ontology*. Edited by David Chalmers, David Manley, Ryan Wasserman. New York: Oxford University Press.
- Schlosser, Markus. (2007). "Basic Deviance Reconsidered." *Analysis*. Vol. 67, No. 3. 186-194.
- Scott, Alwyn. (2007). *The Non-Linear Universe: Chaos, Emergence, Life*. New York: Springer.
- Sehon, Scott. (2005). *Teleological Realism: Mind, Agency, and Explanation*. Cambridge, Mass.: The MIT Press.
- _____. (1997). "Deviant Causal Chains and the Irreducibility of Teleological Explanation." *Pacific Philosophical Quarterly*. Vol. 78. 195-213.

- Sellars, Wilfrid. (1962). "Philosophy and the Scientific Image of Man." In *Frontiers of Science and Philosophy*. Edited by Robert G. Colodny. Pittsburgh, PA: University of Pittsburgh Press. 35-78.
- _____. (1966). "Fatalism and Determinism." In *Freedom and Determinism*. Edited by Keith Lehrer. New York: Random House.
- Shech, Elay. (2018). "Philosophical Issues Concerning Phase Transitions and Anyons: Emergence, Reduction, and Explanatory Fictions." *Erkenntnis*. Published Online (17 January 2018). Accessed December 2018. <https://doi.org/10.1007/s10670-018-9973-z>.
- Shepherd, Joshua. (2014). "The Contours of Control." *Philosophical Studies*. Vol. 170. 395-411.
- Shoemaker, Sydney. (2007). *Physical Realization*. New York: Oxford University Press.
- _____. (2001). "Realization and Mental Causation." In *Physicalism and its Discontents*. Cambridge: Cambridge University Press. 74-98.
- Skiles, Alexander. (2016). "Emergence Reinflated." *Philosophical Quarterly*. Vol. 66. 833-869.
- Stephan, Achim. (2006). "The Dual Role of Emergence in Philosophy of Mind and Cognitive Science." *Synthese*. Vol. 151. 485-498.
- Swenson, Philip. (2016). "Ability, Foreknowledge, and Explanatory Dependence." *Australasian Journal of Philosophy*. Vol. 94, No. 4. 658-671.
- Taylor, Elanor. (2018a). "Only Explanation Can Reinflate Emergence." *The Philosophical Quarterly*. Vol. 68, No. 271. 385-394.
- _____. (2018b). "Against Explanatory Realism." *Philosophical Studies*. Vol. 175, No. 1. 197-219.
- _____. (2015a). "An Explication of Emergence." *Philosophical Studies*. Vol. 172. 653-669.
- _____. (2015b). "Collapsing Emergence." *The Philosophical Quarterly*. Vol. 65, No 261. 732-752.
- Taylor, Richard. (1963). *Metaphysics*. Englewood Cliffs, NJ: Prentice-Hall.
- Temple, Dennis. (1988). "The Contrast Theory of Why-Questions." *Philosophy of Science*. Vol. 55, No. 1. 141-151.
- Thalos, Mariam. (2006). "Nonreductive Physics." *Synthese*. Vol. 149, No. 1. 133-178.
- _____. (2002). "Explanation is a Genus: An Essay on the Varieties of Scientific

- Explanation.” *Synthese*. Vol. 130, No. 3. 317-354.
- _____. (1999a). “Degrees of Freedom: An Essay on Competitions between Micro and Macro in Mechanics.” *Philosophy and Phenomenological Research*. Vol. 59. 1-39.
- _____. (199b). “Degrees of Freedom in the Social World: Towards a Systems Analysis of Decision.” *Journal of Political Philosophy*. Vol. 7. 453-477.
- Tognazzini, Neal A. “Grounding the Luck Objection.” *Australasian Journal of Philosophy*. Vol. 93, No. 1. 127-138.
- Trogon, Kelly. (2013). “An Introduction to Grounding.” In *Varieties of Dependence: Ontological Dependence, Grounding, Supervenience, Response-Dependence*. Edited by Miguel Hoeltje, Benjamin Schnieder, and Alex Steinberg. Munich: Philosophia Verlag. 97-122.
- Varenne, Franck, Pierre Chaigneau, Jean Petitot, and René Doursat. (2015). “Programming the Emergence in Morphogenetically Architected Complex Systems.” *Acta Biotheoretica*. Vol. 63, No. 3. 295-308.
- Walmsley, Joel. (2010). “Emergence and Reduction in Dynamical Cognitive Science.” *New Ideas in Psychology*. Vol. 28. 274-282.
- Wilsch, Tobias. (2016). “The Deductive-Nomological Account of Metaphysical Explanation.” *Australasian Journal of Philosophy*. Vol. 94, No. 1-23.
- _____. (2015). “The Nomological Account of Ground.” *Philosophical Studies*. Vol. 172. 3293-3312.
- Wildman, Nathan. (2018). “Against the Reduction of Modality to Essence.” *Synthese*. Published Online (5 January 2018). Accessed January 2019. <https://doi.org/10.1007/s11229-017-1667-6>.
- Wilson, Alastair. (2018). “Metaphysical Causation,” *Noûs*. Vols. 52, No. 4. 723-751.
- Wilson, Edward O. (1998). *Consilience: The Unity of Knowledge*. New York: Knopf.
- Wilson, Edward O. and Bert Hölldobler. (1988). “Dense Heterarchies and Mass Communication as the Basis of Organization in Ant Colonies.” *Trends in Ecology and Evolution*. Vol. 3. 65-84.
- Wilson, Jessica. (Forthcoming) “Essence and Dependence.” In *Metaphysics, Meaning, and Modality: Themes from Kit Fine*. Edited by Mircea Dumitru. New York: Oxford University Press.

_____. (2015). "Metaphysical Emergence: Weak and Strong." In *Metaphysics in Contemporary Physics*. Poznan Studies in the Philosophy of the Sciences and Humanities. Edited by Tomasz Bigaj and Christian Wuthrich. 251-306.

_____. (2014). "No Work for a Theory of Grounding." *Inquiry: An Interdisciplinary Journal of Philosophy*. Vol. 57, No 5-6. 535-579.

_____. (2013). "Nonlinearity and Metaphysical Emergence." In *Metaphysics and Science*. Edited by Steven Mumford and Mathew Tugby. New York: Oxford University Press. 201-229

_____. (2011). "Non-reductive Realization and the Powers-based Subset Strategy." *The Monist*. Vol. 94, No. 1. 121-154.

_____. (2010). "Non-Reductive Physicalism and Degrees of Freedom." *The British Journal for the Philosophy of Science*. Vol. 61, No. 2. 279-311.

_____. (1999). "How Superduper does a Physicalist Supervenience Need to Be?" *The Philosophical Quarterly*. Vol. 49. 33-52.

Weinberg, Stephan. (1987). "Newtonianism, Reductionism, and the Art of Congressional Testimony." In *Emergence: Contemporary Readings in Science and Philosophy*. Edited by Mark A Bedau and Paul Humphreys. (MIT Press, 2008). 345-457.

Wu, Wayne. (2015). "Experts and Deviants: The Story of Agentive Control." *Philosophy and Phenomenological Research*. Vol. 93, Issue 1. 101-126.

Yates, David. (2016). "Demystifying Emergence." *Ergo: An Open Access Journal of Philosophy*. Vol. 3, No. 31. 809-841.

Zylstra, Justin. (2018). "The Essence of Grounding." *Synthese*. Published Online (31 January 2018). Accessed January 2019. <https://doi.org/10.1007/s11229-018-1701-3>.

VITA

Jordan Baker was born in Memphis, Tennessee (1988) to Tim and Sharron Baker, who homeschooled him and his brother Ethan till college. He attended Union University, pursuing a degree in music theory, and graduated in 2010 with his BM. Jordan then attended University of Tennessee, Knoxville, for a master's in musicology. His master's thesis focused on the relationship between medieval memory practices (*ars memoria*) and the development of Gregorian chant in the 9th -10th centuries. He graduated in 2012 with his MM. Throughout his studies philosophical questions loomed in the background and after completing his master's thesis Jordan decided to have an early mid-life crisis and switch to philosophy for his PhD. He specializes in metaphysics, with particular interests in philosophy of action and metaphysics of nature. He completed the PhD in May of 2019 and looks forward to the next adventure and (hopefully) employment.