

EVENTS AND THE AGENTIAL PERSPECTIVE


Julian Bacharach

A thesis presented for the degree of
Doctor of Philosophy

UCL, Department of Philosophy

In memory of Michael Bacharach

I, Julian Bacharach, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.


.....
Julian Bacharach

IMPACT STATEMENT

This thesis contributes to three principle areas of debate in contemporary philosophy: the philosophy of action, the philosophy of temporal and causal cognition, and the metaphysics of time. In doing so it aims to make some connections between these respective fields and so to advance the terms of each debate.

More broadly, the thesis aims to advance the philosophical project of self-understanding, by making sense of how our lives as agents are related to the world in which we are embedded.

ABSTRACT

There are systematic and pervasive differences between the ways we think about events in our personal past and future. Roughly speaking, we treat past events as fixed and settled, and future events as open and undetermined. This fact raises questions, first, about the finer structure of this pattern of asymmetry; and, secondly, about its metaphysical status. This thesis aims to address these families of question together, thereby bringing questions in the metaphysics of time into contact with ones in the philosophy of action and decision. The principal aim of the thesis is to articulate a distinctively retrospective perspective that we have on our past actions; and to argue that much of mature practical thinking, in particular the asymmetries of past and future thinking, is structured by this retrospective perspective. This perspective is explained in terms of the special epistemic access we have to particular past events, paradigmatically in episodic memory. This is effectively a novel argument for the familiar idea that actions are events—one based on the structure of practical thought, rather than the logical form of action sentences. An ontology of events motivated in this way has consequences for question in the metaphysics of time. Specifically, I argue that it supports a conception of time as a system of particulars, rather than one on which the tenses are fundamental.

ACKNOWLEDGMENTS

One of the claims I argue for in this thesis is that there is a way in which you cannot understand the causes of your action while you are still acting. Although I have no doubt that my way of articulating this thought still leaves much to be desired, nevertheless I am convinced that something like it is true. Reaching the end of a project that has consumed most of my energy and attention for the last several years, I find myself struggling to make sense of what I have actually been up to all this time, and of how I feel about all of it. No doubt it will take me some time to resolve these questions. For now, though, I feel at least knowledgeable enough to recognise, and to thank, those who have guided me through these years of uncertainty.

The greatest thanks are due to my two supervisors, Mike Martin and Rory Madden. Both have been, in different ways, inspirational figures for me. Talking to Mike about philosophy is a uniquely exhilarating intellectual experience. His brilliance lies in getting you to see how the problem that has been bothering you is just the tip of an iceberg; and showing you, with terrifying eloquence, the complex interweaving lines of thought below the surface. But he is also a patient, and profoundly sympathetic, listener and interlocuter; and has an uncanny way of working out what you were trying to say long before you had any idea yourself—in my case, sometimes years before.

Rory meanwhile has been the sober, moderating influence on me. We are, I think, very different kinds of thinker. My tendency is to think everything at once, and then try to work out how it all fits together. Rory has always patiently tried to get me to slow down, separate things out, go one step at a time. Despite my consistent failure to do so, he has never been dismissive, or discouraged me from pursuing my own idiosyncratic lines of thought. He has consistently pushed me to express myself clearly, and been ready with thoughtful, penetrating and insightful comments on my often half-baked ideas. I hope I have not been too exasperating a student.

Last year I visited Berkeley for the Fall 2018 term, and had the opportunity of meeting John Campbell and attending his seminar on causation. *Past, Space, and Self* is

probably the work of recent philosophy that has had the single biggest influence on my ideas, and it was a great pleasure to talk to him about my ideas.

Thanks are also due to Doug Lavin, who at several points read my work, and got me to take seriously the distinctiveness of practical thought. Thanks also to Lucy O'Brien, who supervised me for the MPhilStud, and with whom a conversation in the pub a few months ago enormously helped me clarify the ideas of chapter 4. And to Amia Srinivasan, who took me under her wing as her research assistant, and who many times over the last few years has offered words of wisdom and encouragement.

Perhaps more significant than anything else over the last few years has been the graduate community at UCL, which has consistently been as exciting and welcoming an intellectual home as anything I could have wished for. Amongst former colleagues, three perhaps stand out especially, not just for the many stimulating conversations over the years, but for the impression their intellectual character and virtue has had on me. Vanessa Carr, Alec Hinshelwood and Alex Geddes, in very different ways, exemplify for me a combination of naïve curiosity, and complete seriousness and honesty, that is amongst the best things in philosophy. In this respect I can only try to emulate them. But to say this is in no way to diminish the impact my others peers have had on me. Especially salient to me are Edgar Phillips, Charles Jansen, Dave Olbrich, Laura Silva, Catherine Dale, Pete Faulconbridge, Michael Markunas, Ashley Shaw, Mog Hampson, James Laing, Rowan Mellor, and many others. Thanks are also due to Richard Edwards, for whose tireless work in supporting me and other members of the graduate community I am indebted. And thanks to my great friends, Tom Pye and Will Ghosh, for keeping me on my toes.

It is hard to know how to thank Kirstine la Cour without being hopelessly trite. I am grateful to her for her constant support, understanding and friendship; for listening to me whether or not what I had to say was worth listening to; but most of all, just for being with me, and being herself.

Finally, thanks to my mother, Elizabeth Fricker, who taught me how fun it can be to think.

CONTENTS

Introduction	11
1 The Metaphysics of Perspective and Reality	19
1.1 Perspective, objectivity, and the Absolute Conception	21
1.2 Spatial perspective and sensory perspective	27
1.3 The explanatory gap and the metaphysics of perspective	41
1.4 Abstraction and embedding	45
2 Tense and Temporal Perspective	53
2.1 Tense and truth-conditions	57
2.2 The A-theory and its discontents	63
2.3 Perspective on and in time	74
3 Asymmetries of Causality and Deliberation	81
3.1 Dummett on bringing about the past	84
3.2 Backward causation, causal loops and time travel	90
3.3 Newcomb: Causation and evidence	97
3.4 Intervention and decision	108
3.5 Retrospective Causality	118
4 Regret and Retrospection	123
4.1 Agent-regret, disappointment, and remorse.	124
4.2 Motivation and preference in time	134
4.3 Intention and knowledge of what one is doing	143

4.4	The value of regret	150
5	Temporal Cognition and the Concept of the Past	155
5.1	Temporal frameworks: Script time, phase time, and global time	156
5.2	Understanding causal structure	166
5.3	Memory, agency, and causal self-location	174
6	Tense and Temporal Particularity	179
6.1	Particularity and individuation	181
6.2	Spatial individuation	183
6.3	Tense and temporal particularity	191
6.4	Tense logic and the structure of time	201
7	The Metaphysics of Events	211
7.1	Thinking about events	212
7.2	Existence and quantification	215
7.3	General and singular existence	217
7.4	Substantial existence	224
7.5	Being an event	226
7.6	Events deflated?	229
7.7	Concluding remarks	231
	Bibliography	235

INTRODUCTION

Near the beginning of *Moby Dick*, Melville's protagonist reflects on his decision to sign up for the ill-fated voyage:

Though I cannot tell why it was exactly that those stage managers, the Fates, put me down for this shabby part of a whaling voyage, when others were set down for magnificent parts in high tragedies, and short and easy parts in genteel comedies, and jolly parts in farces—though I cannot tell why this was exactly; yet, now that I recall all the circumstances, I think I can see a little into the springs and motives which being cunningly presented to me under various disguises, induced me to set about performing the part I did, besides cajoling me into the delusion that it was a choice resulting from my own unbiassed freewill and discriminating judgment.

Ishmael's disquiet illustrates two themes I will be concerned with in this thesis: first, the idea of a distinctive causal knowledge one has in relation to one's past actions; secondly, the way in which the location of one's own actions in a wider web of causality raises the spectre of fatalism.

A longstanding Kantian theme is the intimate connection between realism and self-consciousness. Having a conception of an objective world is inescapably connected with a capacity to locate oneself in the world. One variant on this theme, associated with John Campbell's work, stresses in particular that ordinary self-consciousness has an irreducibly causal dimension. The kind of self-awareness exercised in knowing that one is in pain, or in going through the *cogito*, is, according to this line of thought, just one spoke in a nexus of conceptual abilities tied together by a grasp of one's causal structure. Campbell makes a powerful case that this causal dimension to self-consciousness involves a systematic understanding of time as an encompassing causal framework, and in particular involves a realistic attitude to the past, as existing beyond and independ-

ently of its traces in the present.

A question this raises, though, is how a self-conscious grasp of one's causal structure coheres with a conception of oneself as an agent, capable of making a difference. On the one hand it seems obvious that one can only make a difference to the world if one is located in it, plugged into the causal order. On the other hand, there is a difficulty in making sense of this thought from within one's own engaged, first-personal perspective as an agent. One way of bringing this out is to note a connection between objectivity and passivity. In the Third Meditation, Descartes associates the idea of external existence with that of a world 'independent of my will'. And Bernard Williams famously describes a basic condition of objectivity as concerning what is 'there *anyway*'. But practical thought is distinguished in being prior, rather than answerable, to what it concerns; it precisely does not concern what is there anyway. A fully objective conception of the world and one's place in it can then seem to leave no room for practical thought, and so to bring with it the threat of fatalism.

Nowhere is this clearer than in our thinking about time. We have firm realistic intuitions about the past; and this is plausibly connected with a grasp of our past lives as causally embedded in the world. But we largely refrain from applying these same intuitions to the future. How should we make sense of this?

In reflecting philosophically on the difference between the past and future, one can be torn between two extremes. On the one hand, there is the idea that the difference between past and future marks an absolute and objective distinction in reality. This might be articulated in terms of the idea that only the past exists; or that the past is actual but the future merely possible. But this picture is unsatisfactory: it fails to respect the unity and systematicity of our thinking about time, and renders our belief in the reality of the past ultimately mythical. This systematicity in our thinking about time will be a recurrent theme, and in chapter 7, I oppose the recently popular claim that events are confined to the past.

On the other hand, there is the view on which any apparent difference is a projection of habits of thought, themselves explained in terms that presuppose no such objective difference. On the most extreme version of this view, not only the difference between past and future, but also the asymmetry of causality, ultimately drops out from a funda-

mental description of the world and our place in it, and the only temporal asymmetries that remain are *de facto*, non-causal statistical asymmetries. Someone who endorses this view will want to explain the asymmetries of how we act and think in time purely in terms of the availability of evidence to deliberating agents, as a function of asymmetric patterns of mere correlation. I consider this view in chapter 3. There, and in chapter 4, I argue that our use of the notion of cause outstrips its role in guiding deliberation, and hence that asymmetric causal structure cannot be fully explained in terms of the asymmetric structure of evidential inputs to deliberation. This use of cause is involved especially the self-conscious grasp of one's own agency, and specifically in the assessment of past exercises of one's agency.

The broader objection to this second approach is that it entails an intolerable discontinuity between the understanding of time and causality that agents enjoy in acting, and the philosophical account given of the process of acting. The pragmatic, self-centred understanding attributed to the agent's perspective is one on which the only causal relationships are those which obtain in the world outside the agent, and which the agent exploits in order to pursue its goals. An understanding of the agent's own causal structure, by contrast, is one which can be had only from the 'sideways-on' perspective of the theorist. This mistake is, in the end, not so different from the mistake involved in taking the difference between past and future to be metaphysically primitive. Both approaches fail to recognise the respect in which self-conscious agents, in acting, deploy a systematic and objective understanding of the order of time and causality, and of themselves as located within that order.

Identifying a mistake does not make avoiding it straightforward. The difficulty is precisely to make sense of how the engaged, active perspective of an agent involves, fundamentally, both a temporal orientation in which the future is taken to be open and the past fixed; and at the same time a systematic appreciation of the structure of time as a causal structure in which one is oneself enmeshed. This difficulty is not just a philosopher's invention: it is one that can strike us, as it does Melville's protagonist, in the midst of ordinary self-conscious reflection, when we step back and try to make sense of the overall shape of our lives. Discovering the hidden springs of one's action can, on occasion, produce something like the sense of having been set up; an insight into the

causal structure in which one is operating can destabilise one's sense of being in control.

Acknowledging this tension can tempt one to posit two levels in our thinking about time: a 'hot', engaged, agential level, in which the agent themselves is absent, or occurs only as a vanishing point, in the represented causal order; and a 'cool', detached level, at which the agent reflects on the structure of time and their place in it. I want to argue that this thought is half right. It is right insofar as adopting the stance of a deliberating agent sometimes prevents one from knowing certain things about the causal structure of one's situation, and, in particular, prevents one from reflecting on the causes of one's actions. I shall argue in chapter 3 that this is the real lesson of Newcomb's problem; and in chapter 4, I shall argue more generally that there are causal insights into the histories of one's actions which can only be had in relation to the past.

To this extent, there is a difference between the immersed perspective of deliberation, and the reflective project of causal self-location. Our position as agents imposes a limitation on how much we can know about the causal structure of our agency in the present and future. However, I want to resist the move from this observation to the idea of two fundamentally different levels of thought, one agential and one that is fundamentally disengaged from one's nature as an agent. Objective thinking about time and causality is involved in practical thought as much as in detached scientific theorising. We think objectively about time and our place in it whenever we reflect on our past actions, and try to make narrative sense of our lives. This kind of thinking is reflective, but not disengaged. Mature ethical thought cannot get by without a sense of a wider objective order, beyond our immediate environment and appetites, in which we have our place. The point is that, due to our fundamental orientation as agents, we apply this objective conception differentially with respect to the past and future.

In the latter part of the thesis I argue that this difference between past and future thinking is to be articulated in terms of a temporally differentiated capacity to single out particular events, and employ singular causal explanations. I argue, in chapters 5 and 6, that the ability to engage in historical causal thinking involves possession of a global temporal framework; and that possession of the temporal framework is supported by a capacity to single out particular events. This introduces an importantly perspectival element into our understanding of time: our command and use of the global temporal

framework depends critically on our ability to single out and locate the particular events that make up our lives as agents.

Throughout the thesis, I adopt a critical though receptive attitude to the claims of metaphysics. I shall be interested in questions of the reality of the manifest world as it appears to us, and in particular how our manifest image of time relates to temporal reality. This includes, for instance, the question whether there is a real difference between the past and future, or whether any apparent such difference is just the shadow or reflection of our habits of thought. My attitude is critical in that I think these questions have to be understood as, at bottom, questions about the nature of the perspective we occupy on time. But I am receptive to metaphysics in that I do not think that our perspective can be adequately characterised without making assumptions, or incurring commitments, about what the world is actually like. I shall argue that our temporal perspective is complex; and that, to understand its complexity, we have to see it as arising from, on the one hand, our embeddedness as agents in a world of causes; and, on the other hand, the way in which as self-conscious agents we conceptualise our causal connectedness to the surrounding world.

CHAPTER SUMMARIES

CHAPTER 1 The first two chapters constitute a kind of extended introduction to the issues of time and agency discussed in the rest of the thesis. Chapter 1 sets up a background metaphysical dialectic concerning the question whether a given phenomenon is real, or whether a certain area of our thought reflects reality. Drawing on Bernard Williams's influential discussion of the 'absolute conception of reality', I examine the frequently encountered theme that perspective is a mark of unreality: that certain of our beliefs and concepts, because they are marked by idiosyncrasies of our perspective, therefore fail to represent reality perspicuously. I argue that the paradigmatic case of this kind of perspective in our representations is the use of egocentric spatial notions, such as 'left' and 'right'. However, I argue that the extension of this model beyond the paradigmatic spatial case, for instance to the case of colours, is often obscure and unhelpful. Against this, I contrast a different way in which our representations may be said

to depend on our perspective: the sense in which our perspective is our particular and contingent route through the world, and the content of our representations is explained by our relations to the objects and features we encounter. I call this conception of our relation to the world the ‘embedding conception’. I suggest that the embedding conception offers a fruitful way of making sense of our relation to the world in general; and, specifically, of the dependence of particularity in thought on our contingent specific relations to particular things in the world.

CHAPTER 2 This chapter extends the dialectic of perspective and reality from the previous chapter to consider specifically the question of the reality of tense. I argue that anyone who wants to maintain the reality of an absolute difference between past and future needs to contend with the ways in which our understanding of time is systematic, and involves an appreciation that the past is not an independent and autonomous domain. At the same time, we need to recognise that temporal experience does not seem to present us with a unified domain encompassing a plurality of perspectives in the way that spatial experience does. I suggest that the embedding strategy may provide a way to overcome this impasse, by recognising the way in which our conceptual command of the temporal domain is interdependent with our actual causal embeddedness in structures and processes with an intrinsic temporal orientation. This abstract suggestion is developed and implemented in the subsequent chapters.

CHAPTER 3 This chapter concerns the basis of the fact that we deliberate over the past and not over the future. An obvious explanation is that causes are ways of bringing about their effects, and not vice versa; and causes always precede their effects. However this pragmatist idea of causes as ‘handles’ becomes problematic in cases where our actions are not purely exogenous inputs to the environment we are acting on, but may have common causes with whatever result we are aiming to bring about. I discuss such cases in the archetypal setting of Newcomb’s problem, and the ensuing debate between Causal and Evidential Decision Theory. I argue that the causalist’s recommendation cannot be justified on purely pragmatic grounds, and that we need to appeal to the retrospective evaluation of one’s decision in order to justify the relevance of causal structure to deliberation. This is an instance of the more general idea, to be developed in later chapters,

of a distinctively retrospective, explanatory use of causal concepts, involved in a grasp of one's causal structure as an agent.

CHAPTER 4 This chapter further develops the idea of a retrospective perspective of one's own actions, drawing on Bernard Williams's discussion in 'Moral Luck' of agent-regret. I argue that such retrospective attitudes should fundamentally be understood in terms of the role they play in processes of motivation and action. I compare regret with relief, a different attitude that possesses an intrinsic temporal orientation. In the case of relief, I argue that its temporal profile should be understood in terms of its connection with dynamic processes of motivation, and in particular with the reaction of aversion. I then argue that intention, and intentional action can be understood as having essentially the same dynamical character, and that this makes room for the idea of thoughts about one's own action that are tied to specific points in motivational processes. Broadly speaking, the idea is that acting intentionally means taking an interested, engaged attitude to what one is doing; and that this attitude precludes a certain kind of reflection, associated with an understanding of the specific psychological causes of one's actions. This makes way for the general idea that there is a certain kind of reflective understanding of oneself as an agent, and one's causal embeddedness in the world, that is essentially retrospective.

CHAPTER 5 In this chapter I consider how the kind of retrospective causal thinking discussed in the previous two chapters is supported by our ways of thinking about time. I argue that historical causal thinking requires an understanding of events as located within a global, connected temporal framework. I contrast this with some less sophisticated ways of representing time, also discussed by John Campbell, and by Christoph Hoerl and Teresa McCormack. I argue that possession of a global temporal framework is connected specifically with the ability to think causally about oneself—precisely the kind of causal thinking that is excluded from the pragmatist conception of cause discussed in chapter 3.

CHAPTER 6 This chapter further explores the idea that our thought about time concerns temporal particulars. Drawing on an early discussion of Russell's, I argue that the

notion of a particular is intimately associated with the distinction between qualitative and numerical identity. I argue that, in the case of spatial particulars, our ability to distinguish qualitatively similar particulars on the basis of their relations to one another rests on the intrinsic particularity of spatial experience, as well as conceptual command of the spatial framework. I argue that the same must be true in the temporal case—that temporal experience fundamentally concerns particular events.

CHAPTER 7 This chapter addresses a recently popular claim that events exist only once they are over. If true, this claim would be problematic for the idea that temporal experience intrinsically concerns particular events. However, as I argue, the claim is incoherent. I consider two ways of making the claim, and the notion of existence it involves, more precise: one that ties existence to the logic of quantification, and another that assimilates the existence of an event to that of a continuant. The first of these I argue cannot be the right way to understand the target claim, since it leads to an outright contradiction when applied to claims of singular (temporary) non-existence. The second, on the other hand, leaves it ultimately mysterious what the past existence of an event could amount to. I conclude that any credible ontology of events must take them to exist in the present, as well as the past.

1 | THE METAPHYSICS OF PERSPECTIVE AND REALITY

A very general style of metaphysical question is to what extent, and in what respects, one aspect or other of our understanding or experience of the world matches up with the world itself, with reality. This question may, and has been, posed with regard to many diverse areas of our thought: for ethical and evaluative discourse, for thought and experience of colour, for numerical thought, for beliefs about necessity and causality, and so on. For each of these aspects of our thought it is easy to be subject to the anxiety that what we are doing when we think in this way is somehow more, or less, than just faithfully and transparently representing what is ‘out there’; that we represent the world in ways which are partial, or distorted, or embellished.

This anxiety is closely connected with the recognition that our encounters with the world are always from some particular, idiosyncratic perspective. We want to be able to sort out those features of our worldview which are due to the peculiarities of the perspective we occupy from those which reflect how things really are. The metaphysical question can be seen as the most general and systematic expression of this process of reflective correction.

Another such domain which the metaphysical anxiety can fix upon is our experience of the temporal world. A central feature of our understanding of time is the sense that time passes. It is obscure what exactly this sense of passage amounts to, but at least one aspect of it is this: first, there is a real difference between the past and the future, one which might be roughly expressed by saying that the past is fixed and settled, whereas the future is open and indeterminate; secondly, what is past and what is future changes.

We might try to express this change by saying either that there is a forward-moving ‘now’, a frontier which advances from the past into the future; or perhaps that events move from the future into the present, and recede into the past. But it is doubtful how illuminating either of these metaphors of motion can be, since the very notion of motion as opposed to stasis is one that presupposes the passage of time.

The version of the metaphysical anxiety that reflection on the passage of time gives rise to is often formulated as a question about the status of tensed representations—representations that classify events as past, present, or future. This threefold division of events is what J. M. E. McTaggart famously named the ‘A-series’, in contrast with the ‘B-series’, which orders events by a binary relation of earlier-later. When the A-series is superimposed on the B-series, the result is an ordering of events centred on a unique present time, with earlier times classified as past and later times as future. (For now I am passing over questions of whether the relevant domain is one of concrete events, instantaneous times, or temporal intervals.) We can then ask the question whether this division of events, and the identification of a unique present, corresponds to anything in reality, or whether temporal reality in itself is untensed.

This is the question of the reality of tense. In one guise or another it is at the centre of most metaphysics of time in the Anglophone tradition since McTaggart’s discussion. But the question is highly elusive. What is at issue is not whether tensed representations are necessarily false or inaccurate: everyone agrees that tensed representations can be true, so it is not straightforwardly a matter of whether, in using tensed representations, we are thereby attributing to the world some spurious feature. Nor is it obviously a question of whether a complete and ideally perspicuous description of reality would involve only tenseless representations. Given that reality contains particulars which are discernible only by reference to their spatiotemporal position, it may be that any complete description of reality must perforce employ indexical or self-locating vocabulary in order to single out one particular rather than a qualitative duplicate. This would witness the necessity of describing reality from some perspective or other, but would not show that any such perspective was intrinsic to the reality so described.¹ Just so, it might be that tense is ineliminable from any complete description of reality just because reality

¹This point is borrowed from Campbell 1993, and goes back to P. F. Strawson’s *Individuals*.

can only be completely described from some temporal perspective or other, and tense is a mark of temporal perspective. But this would not seem to indicate that tense, or the temporal perspective it expresses, is itself a feature of reality.

The question is meant to be whether reality is *itself* tensed; whether tense is an intrinsic feature of reality. But the notion of tense understood as a feature of reality, rather than of the representations which describe it, is ultimately obscure, and necessarily so. We have, in effect, introduced the notion as a semantic feature of a certain fragment of discourse, namely temporal discourse; but in asking whether this semantic feature ‘answers to’ anything in reality, we have to prescind from language and form some conception of what it would be for reality to contain tense, independently of its being represented (and truly represented) in tensed terms. This obscurity does not necessarily vitiate the metaphysical question; but it does signal a need to be alert to the basis and limits of our grip on the notions in play.

In the rest of this chapter I will sketch a general dialectic that issues from the metaphysical question of appearance and reality, focusing principally on the more familiar case of colour. This will enable us to get a better grip on the various available ways of understanding the key notions of perspective, fact, and reality, and their interrelations.

1.1 PERSPECTIVE, OBJECTIVITY, AND THE ABSOLUTE CONCEPTION

The general form of the metaphysical question is: To what extent does an aspect of our experience or understanding of the world reflect reality? Assuming that the relevant aspects can be made explicit, and given discursive articulation, this question can be rephrased as one about language or discourse: To what extent does a given fragment of discourse correspond to reality? The question can then be given a more focused interpretation by isolating some bit of a theory—a set of sentences, or body of beliefs—and asking about whether the relevant vocabulary, or repertoire of concepts, in which the theory is couched pick out anything in reality. For example, in the case of the metaphysical question about colour, at the most general level of framing it is helpful to abstract from the details of colour experience, and ask whether our body of beliefs about colour, or our use of colour concepts, faithfully reflect the facts about objects and their colours.

The step from a more general, nebulous notion of ‘how things seem’ as the starting-point for the metaphysical question to a more focused, theory-driven one is in fact not trivial. It assumes that the relevant aspects of how things seem are sufficiently systematic and coherent that they can be brought to articulacy in the right way without introducing distortions of theory. Later on I shall argue that, in the case of how things seem with the temporal world, it is precisely this assumption that is problematic. However for the present purpose of exploring the general shape of the metaphysical question, I shall follow this step and merely note the assumption on which it depends.

The distinctive force of the metaphysical question comes from the fact that it is not normally in dispute whether the relevant fragment of discourse is, for the most part, *true*, at least by ordinary standards. Of course judgments of colour are fallible, and it might be that there are certain ways in which humans are prone systematically to make mistakes about the colours of things.² But the idea is rather that, even when we get it right, we are nevertheless not getting onto reality. The difference between a true and a false judgment about colour does not correspond to any significant difference in how things really are.

This thought is given expression in a famous discussion by Bernard Williams:

We can draw distinctions between things seeming green and their really being green...But these usages do not go very deep...Our distinctions between what seems green and what is green are essentially based on agreement within the range of human experience, and human thought is not, in that limited sense at least, tied only to human experience: scientific and philosophical reflection can stand back from at least these peculiarities of our constitution. (1978, p. 242)

The sense in which the difference between being green and seeming green, or between a true and a false judgment that something is green, does not go very deep, is that when we

²Barry Stroud puts the point like this: ‘We can certainly have the coherent thought that part of what we believe about the world might not be true, and that for peculiarly human reasons we easily fall into error or illusion in certain characteristic ways. But the thing to do when that possibility is raised is to re-examine our beliefs and to be especially vigilant in looking for partiality, self-interest, peculiar perspective, limited experience, or the influence of other “human peculiarities” in our reasons for holding them. There is no end to the task of getting the best supported set of beliefs we can; at every point we must admit that error or illusion might still remain undetected. But the search for an “absolute” conception of the world is supposed to begin only after such epistemic questions have been answered as well as they can be for the moment. It is not simply an epistemic search for what to believe.’ (2000, p. 40) My understanding of the matters discussed in the following pages owes a lot to Stroud’s discussion.

do stand back from the ‘peculiarities of our constitution’ we arrive at a conception of the world that does not, fundamentally, involve colours. The paradigm of this conception of the world is of course that provided by physics.

There are two interlocking components to this line of thought. The first is that some of the judgments we make, or the distinctions we draw, betray idiosyncratic aspects of our epistemic or experiential perspective on the world. In the case of colour, this would be the fact that we are equipped with a sensory apparatus that responds in highly specific ways to different wavelengths of light. The second component involves putting forward, in however schematic terms, a positive conception of reality, purged of those idiosyncratic perspectival features, which does not attribute colour to things. Providing this positive conception is supposed to make good on the initial suspicion that distinctions of colour were just an artefact of perspective, rather than identifying a real difference in the world.

This process, whereby some area of our thought is suspected of being due merely to peculiarities of our constitution, and is then supplanted by a more austere conception from which the problematic concepts are absent, is described by Thomas Nagel:

The first step is to see that our perceptions are caused by the action of things on us, through their effects on our bodies, which are themselves part of the physical world. The next step is to realize that since the same physical properties that cause perceptions in us through our bodies also produce different effects on other physical things and can exist without causing any perceptions at all, their true nature must be detachable from their perceptual appearance and need not resemble it. The third step is to try to form a conception of that true nature independent of its appearance either to us or to other types of perceivers. (1986, p. 14)

Once again, the paradigm of this progress of objectivity is supposed to be scientific enquiry, which abstracts away as far as possible from the peculiarities and limitations of human observers. Both Williams’ and Nagel’s characterisations of natural science as playing this role are largely modelled on early modern conceptions of the enterprise, and one might perhaps question the extent to which the contemporary sciences, with their vast plurality of concepts and methods, fit this mould. Nevertheless it is still true at least of contemporary physics—and perhaps more relevantly, vision science—that its theories are cast in mathematical and informational terms which in some sense purport

to be less infected with the peculiarities of a distinctively human perspective.

The picture of colours and other qualities as receding from the world conceived scientifically, in abstraction from the human perspective, is undeniably compelling. But there are pressing questions to be asked about it. One question is: To what extent does the fact that colours do not show up in a purely physical description of the world, or in the descriptions of human perceptual processing offered by vision science, license the conclusion that colours are not ‘really’ part of the world, that colour discourse does not correspond to reality? It is easy to slide from the notion of reality, or ‘the world as it really is’, to ‘the world as described by physics’. But even if we take it as given that everything in reality physical, this is consistent with many physical things having non-physical characteristics. On the other hand, if we just identify reality with the totality is what is asserted by statements couched in purely physical vocabulary, then, on the assumption that colour concepts are not part of physics, it follows trivially that statements about colour do not describe reality. But some reason needs to be offered for this identification.³

A second, related question is exactly how to make sense of the idea that the acknowledged truth of most colour discourse is due to features of our perspective, specifically to the constitution of human perceivers. Neither Nagel nor Williams wish to endorse the crass subjectivist claim that things are only coloured when they are perceived, or that things would not be coloured if there were no colour-sighted perceivers.⁴ So it is not that things’ being coloured crudely depends on the operations of human sensibility, nor that human sensibility leads us into error by having us falsely attribute colours to things. Rather the idea is that the level of description at which things can be (truly) said to have colours is one formed by, and bearing traces of, the human perspective. But the notion of a perspective, one which may be more or less idiosyncratic or subjective, still needs to be made clear.

The questions are related because it is plausible to expect that a satisfactory answer to one should provide materials for an answer to the other. What the first question demands is why we should treat the austere scientific conception as authoritative about what is really there, rather than just as incomplete. Such a reason will presumably in-

³Cf. Stroud (op. cit.), pp. 61–65

⁴Williams, *ibid.*, p. 242

volve the idea that the scientific conception is offered from a perspective which is less parochial, and more objective, than everyday conceptions which involve such things as ordinary objects and their colours. And so this in turn should explain the sense in which colour discourse depends on the perspective of human subjectivity.

The key thought here is that the scientific conception is meant not just to replace the various more parochial conceptions which contained such things as statements about colours, but to subsume or transcend those conceptions. It will be in a position to do this if it has the resources to accommodate, within the world it describes, the existence of the perspectives to which the conceptions it subsumes belongs, and to explain how certain features of that perspective contribute to how things appear to someone who occupies it. The modern conception of colour experience, alluded to by Nagel, as generated by purely mechanistic interactions between physical bodies, is offered as a prime example of this subsuming operation. The authority of the mechanistic conception of nature resides in its putative ability to explain, in its own proper vocabulary, why it is that things appear coloured to perceivers with the appropriate sensory apparatus, in normal lighting conditions, and so on. The crucial idea is that the account of the perspective can be given in terms that do not require one actually to occupy it in order to understand them. For example, a mechanistic account of the operation of human colour perception is supposed to be in principle available to creatures who do not perceive things as coloured. In offering such an explanation, the everyday conception of things as coloured is understood to belong just to the perspective of perceivers with that kind of sensory apparatus; this provides the sense in which the perspective of colour discourse is distinctively human or subjective.

Williams envisages an ideal limit of this process, which he calls the ‘absolute conception of reality’: a maximally neutral, perspective-free representation of how things really are, extended to explain the existence of the various local perspectives and how things appear from them, including the absolute conception itself. The core of the absolute conception is, presumably, a mathematical physics. But, as Williams recognises, this is not enough: if the human perspective is to be incorporated into the absolute conception, we need to supplement pure physics with a science of human nature and activity: not just vision science, but psychology, history, social science, and much more.

This, however, raises an urgent question about the authority of the absolute conception. Mathematical physics alone manifestly does not contain materials sufficient to explain the existence of the more specific perspectives which shape our everyday engagement with the world. So in what sense can we legitimately regard those more everyday conception as having been subsumed or transcended?

Williams is well aware of this tension, writing,

The substance of the absolute conception...lies in the idea that it could non-vacuously explain how it itself, and the various perspectival views of the world, are possible. It is an important features of modern science that it contributes to explaining how creatures with our origins and characteristics can understand a world with properties that this same science ascribes to the world. The achievements of evolutionary biology and the neurological sciences are substantive in these respects, and their notions of explanation are not vacuous. It is true, however, that such explanations cannot themselves operate entirely at the level of the absolute conception, because what they have to explain are psychological and social phenomena, such as beliefs and theories and conceptions of the world, and there may be little reason to suppose that they, in turn, could be adequately characterized in nonperspectival terms. (1985, pp. 139–140; cf. also 1978, pp. 301–302.)

He goes on to suggest that, even if a perspectival element is ineliminable from our understanding of the perspective of human knowledge, this does not undermine the value of the absolute conception, because ‘It will be a conception of nonperspectival materials available to any adequate investigator, of whatever constitution, and it will also help to explain to us, though not necessarily to those alien investigators, such things as our capacity to grasp that conception.’ (Ibid.)

This last point, however, is not entirely satisfactory. It is certainly true that the conception of the world provided by physics has great generality and explanatory power, in the sense that physical principles apply universally and without exception throughout the universe. This is an important contrast with, say, the kind of physics envisaged by Aristotle, where the behaviour of each thing is explained by special principles proper to its specific nature. (Although again, perhaps the early modern conception of a universal mechanics is closer to the ideal here than the rather more fragmented picture that emerges from contemporary physics.) But generality alone is not a sufficient basis for the claim of the physical conception of the world to transcend other, more local conceptions. Those other conceptions—for example, any conception which characterises

ordinary objects as coloured—were revealed to depend on idiosyncrasies of perspective precisely because of the possibility of giving an account of that perspective, and of how things appear to someone occupying it, which could be understood without occupying that same perspective. Without this kind of explanation on offer, it looks as though the different conceptions may just be concerned with different aspects or levels of reality, the main difference being that the physical aspects of reality are more systematic and pervasive than others. There may not be any interesting sense in which the more general, physical conception transcends the more local ones. Thus if the social and psychological explanations required to bolster the absolute conception cannot, as Williams suspects, be ‘adequately characterized in nonperspectival terms’, we lose our grip on the sense in which those explanations are indeed any more perspectival than those proper to physics.

The problem is that the notion of a perspective being revealed as idiosyncratic in being explained by, and so subsumed within, some more comprehensive theory, remains obscure. But this subsumption operation is crucial to the progress of objectivity described by Williams and Nagel, and so to the metaphysical question whether reality corresponds to how things seem. We should take a closer look at what is going on.

1.2 SPATIAL PERSPECTIVE AND SENSORY PERSPECTIVE

In a crucial passage, Williams writes:

Suppose A and B each claims to have some knowledge of the world. Each has some beliefs, and moreover has experiences of the world, and ways of conceptualizing it, which have given rise to those beliefs and are expressed in them: let us call all of this together his *representation* of the world (or part of the world). Now with respect to their supposed pieces of knowledge, A's and B's representations may well differ. If what they both have is knowledge, then it seems to follow that there must be some coherent way of understanding why these representations differ, and how they are related to one another. One very primitive example of this would be that A and B were in different places; another might be that they were both correctly predicting the movements of the planets, but by different, geometrically equivalent, systems. In either case, a story can be told which explains how A's and B's can each be perspectives on the same reality. *To understand this story, one needs to form a conception of the world which contains A and B and their representations;* A and B are not debarred from taking this standpoint themselves, but it involves

their standing back from their original ways of representing these aspects of the world. (op. cit., pp. 64–65; emphasis added.)

Williams presents this process, of integrating different perspectival representations into a more comprehensive picture of the world, as the first step on the path to the absolute conception. In this way he argues that the possibility of the absolute conception is implicit in the very idea of knowledge.⁵

Two remarks are worth making here. First, there are two characters in this story, A and B: we are given a multiplicity of perspectives, and feel a need to integrate them. This is importantly different from the case of colour. There, we start from a position of widespread agreement about the colours of things, but are prone to an inchoate sense that there is something less than fully real about them. Of course things are different with other matters—when it comes to questions of morality and value, we do seem to be confronted with genuine differences that demand reconciliation, or at least explanation. Still, it is not yet clear what the connection is between this multiplicity, and the sense of unreality that prompts the metaphysical question.

This leads on to the second point about the above passage. Williams is careful to say merely that A and B's representations differ, whilst being noncommittal about what this difference comes to. In particular, he does not stipulate that it is in the nature of a *disagreement*. If it is a genuine disagreement then, more or less by definition, only one of them is correct, and so there is no possibility of integration into a single conception—the problem is rather one of arbitration. On the other hand, there may be ways in which A and B's representations 'differ' that do not create even any *prima facie* conflict. Perhaps A knows a lot about tropical plants and B knows a lot about Urdu poetry. Then their representations can be 'integrated' just by conjunction, pooling their expertise in their respective fields into a single body of knowledge. The resulting theory will be more comprehensive insofar as it contains all of A's and B's representations as parts; but it is unclear what the point would be of saying that the perspectival character of either's knowledge has been revealed.

The canonical case Williams has in mind is, perhaps, one in which there is an *apparent* disagreement between A and B, which is resolved by forming a more comprehensive

⁵An detailed expansion of this argument is presented in A. W. Moore 1997, ch. 4.

conception that incorporates their representations. Perhaps A thinks that there is a bright star to her left and B thinks there is one to his right. Clearly, they can both be correct, in two main ways: either they are not looking at the same star, or they are looking at the same star but facing in opposite directions. In either case, the disagreement is revealed as merely apparent. The first possibility is a version of the case where A and B are experts in different fields: A is an ‘expert’ on the location of her star, B on his. The second possibility, on the other hand, requires the recognition that left and right are not incompatible qualities: that being on the left or the right depends on which way one is facing.

It is worth dwelling on the similarities and differences between these two ways in which the apparent disagreement may be resolved. Both involve the recognition that how things seem to one depends upon one’s perspective—specifically one’s spatial position. In the case where A and B are looking at two different stars, integration just requires the recognition, not always trivial, that they are indeed talking about different things. In this case the relevant point is that what one sees depends upon where one is looking, and hence two differently situated observers may fail to disagree just because they are observing different scenes. The notion of perspective introduced here is just the idea that which bits of the world a thinker apprehends—which objects they perceive, or which truths they know—depends upon their particular position in and route through the world. A thinker’s perspective in this sense may be idiosyncratic in that it is local, restricted to a particular portion of the world in virtue of the thinker’s particular position or constitution. But notice this does nothing to undermine the status of the thinker’s representations as reflecting reality.

The second case, on the other hand, requires a distinctively *semantic* insight into how ‘left’ and ‘right’ work, namely that whether it is correct to say that something is on the left or the right depends upon one’s spatial position and orientation. This insight is crucial: it embodies the recognition that the same bit of reality may be presented and described in different ways. Coming to appreciate this feature does seem tantamount to the recognition that ‘left’ and ‘right’ do not, on their own, represent features of reality: rather they express different ways of thinking of the same bits of reality.

We thus have three fundamentally different models for how we can ‘make sense of’

differences between different observers' perspectives. There is the case of genuine disagreement, in which there is no possibility of integration. There is the case where the perspectives are on two different bits of reality, and therefore the observations consistent. And there is the case where the two observations are consistent because they are describing the same bit of reality in different ways. In the first two cases, we may talk of the thinker's perspective in the sense of those features of their contingent relations to the world that contribute to their having the representations they do—spatiotemporal location, sensory apparatus, cultural upbringing, and so on. This is a perfectly legitimate notion of perspective, but one which does not deliver any sense in which the thinker's representations, even if true, do not reflect how things are in reality. It is only in the third case, where we have multiple perspectives on the same bit of reality, that this verdict becomes a salient option. We should pursue this thought further.

The phrase 'same bit of reality' here has a special sense. The relevant case is not one where thinkers make different predications of the same object: one person thinks the apple is green, the other thinks that it is round. That case is just another variant on the one where A and B are looking at different stars; that the same individual features in both representations here is irrelevant. The notion of a bit of reality in play here is rather that of a fact or state of affairs: something with quasi-predicational complexity mirroring that of a sentence or proposition.

The most minimal possible notion of a fact is purely disquotational. We might think of this notion of a fact as introduced by a redundant sentence-level operator, 'It is a fact that _____', or as a predicate of sentences satisfying the schema 'sentence S states a fact iff φ ', where ' φ ' is always a translation of S. Clearly this notion of a fact will not provide us with any interesting sense in which two sentences that are not mutual translations, or two different representations, can state the same fact. It is not that this conception of a fact requires a one-one correspondence between true sentences (or classes of translationally equivalent sentences) and facts; indeed it is consistent with there being only one single fact. Rather the point is it gives no way of determining the identity conditions of facts, so provides no substance to the idea that some pairs of true representations represent different facts and other pairs represent the same fact.

A more promising, and slightly less minimal, notion of fact ties their identities

to (true) propositions, or to the truth-conditions of sentences and other representations. This will be an advance if we can make out an interesting sense in which different representations—or more precisely, representations of different types—can ‘say the same thing’: share truth-conditions, or express the same proposition.⁶

This condition does indeed seem to be provided for in the case of sentences whose evaluation as true or false is sensitive to the context of use: sentences involving terms like ‘I’, ‘here’, ‘this’, or—more controversially, in the light of the discussion of the following chapter—‘now’. The natural thing to say about such expressions is that the truth-conditions of sentences containing them are determined not by their components alone, but by their components in combination with relevant features of the context of use. The same sentence used in two different contexts can thus determine two different truth-conditions. Perhaps slightly more controversially, two different sentences used in different contexts can determine the same truth-condition.

This is a plausible story about how, when A says ‘A bright star is on my left’ and B says ‘A bright star is on my right’, they can both be stating the same fact. An utterance involving ‘left’ or ‘right’ requires completion with two contextual parameters in order to determine a truth-condition, namely a position and a direction of orientation. The sentences ‘X is on the left’ and ‘X is on the right’, when supplemented with (roughly) the same position and opposite directions of orientation as contextual parameters, determine the same truth-condition. We can think of these two parameters as constituting the perspective of any representation involving the concepts ‘left’ or ‘right’. Such representations are perspectival in the sense that their truth-conditions depend on the contextual parameters which constitute the perspective from which they are made.

To what extent does this simple semantic story deliver the conclusion that left and right are not features of reality, and that representations couched in terms of ‘left’ and ‘right’ do not represent things as they really are? One might think not very far. What it establishes is that ‘left’ and ‘right’ do not express monadic properties which an object can be said to have or not have, absent further relativisation. But the natural conclusion

⁶This of course raises the question whether, for instance ‘Hesperus is bright’ and ‘Phosphorus is bright’ state the same fact, since, on some plausible criteria for same-saying, they do not say the same thing. One may therefore want to allow a notion of fact coarser than that of a proposition in allowing sentences with co-referential singular terms (occurring in extensional contexts) to state the same fact.

to draw is surely not that left and right are not real features, but that they are three-place relations between an object, a point of reference, and an orientation vector.⁷ This much is, plausibly, embedded in our ordinary understanding of how the terms ‘left’ and ‘right’ work: we say, not just that an object is on the left or on the right, but that it is to *my* left and to *your* right, and we understand that both of these may be true if you and I are facing different directions.

We can however imagine a thinker who has not fully grasped the relational nature of left and right. A creature might operate purely with egocentric spatial representations in which left and right are used as monadic spatial dimensions, along with forwards, backwards, up and down.⁸ If the creature is mobile, it would need to continually update its representations in terms of these three axes in order successfully to navigate its environment. But successful updating does not necessarily require any kind of superordinate representation in which left and right, forward and back, are represented as relations. We might well say that this creature does not yet have a proper grip on spatial reality, that their spatial notions do not properly represent how things really are spatially organised—notwithstanding the fact that, in some weaker sense, it is capable of successfully and accurately representing the locations of environmental features.

We can understand the kind of revision involved here by reflecting on cases where the relativisation to a further parameter is somewhat less obvious. Places are ordinarily individuated and reidentified in relation to an array of objects serving as a frame of reference, and hence attributions of motion and velocity involve an implicit relativisation to that frame. Sometimes this relativisation becomes explicit: it might be appropriate to say, of a sleeping passenger on a train, both that they have been stationary for the last hour, and that they have travelled 80 miles in the last hour. Clarifying the sense of either statement would involve specifying whether one was speaking relative to the interior of the train, or to the earth’s surface. The need to make a frame of reference

⁷It is important to distinguish the sense in which left and right are relational properties from a closely related sense in which they are intrinsic: the sense in which a person has a left and right hand, and there is an intrinsic difference between a left- and right-handed glove. Left- and right- orientation in this sense are part of the intrinsic geometry of an object, related to the distinction between clockwise and anticlockwise rotation. The direction ‘to the left’ is equivalent to (though not necessarily defined as) being on the left-hand side of the sagittal plane; that is, the same side of the sagittal plane as the left part of one’s body. This of course depends upon the functional differentiation of the body into front, back, top and bottom.

⁸For ‘left’ and ‘right’ as monadic predicates, see Campbell 1994, ch. 1.

explicit can arise when the objects of our discourse can be grouped differently so as to provide multiple frames of reference which are moving with respect to one another—as when speaking of the interiors of transportation devices. People living in less technological societies might never encounter these situations. They might then view the landmarks by which they navigate their environment as cues to an underlying framework of absolute spatial positions, rather than as means for designating a useful but arbitrary standard spatial frame of reference. This is, arguably, how we technologically advanced people unreflectively think of the role of environmental landmarks. Unlike with ‘left’ and ‘right’, it is not obviously part of basic competence with the concepts ‘same place’ or ‘moving fast’ that one understand these expressions to carry an implicit relativisation to a contextual parameter.⁹

There does seem to be a genuine sense in which a community of speakers who use monadic predicates of rest and motion without any acknowledgment, however sketchy or inarticulate, of the need to relativise to a frame of reference, are failing to represent reality perspicuously. Their discourse or conceptual scheme makes an absolute distinction between rest and motion, whereas in reality there is no such distinction. (Or if there is, it does not map onto theirs: the only way to save their discourse from simple falsity is to interpret it as requiring relativisation to their local frame of reference—see n. 9 above.) This point could be dramatised by supposing a second community who are moving with respect to the first. Imagine two communities on respective planets of a small binary system, each thinking their planet is at rest and the other in orbit. Perhaps their astronomers calculate the movements of the heavenly bodies in ways that appear to be at odds, but both always yield correct predictions. Someone might be tempted to say that the astronomers are living in ‘different worlds’, working in incommensurable scientific paradigms. But this would clearly be a feeble cop-out. Conversation between

⁹Note that this point is distinct from the question whether there is an absolute metaphysical distinction between rest and motion. That question turns principally on how physical theory connects purely kinematic differences with differences in observable forces, and on related modal considerations about whether a given theory of spacetime spuriously multiplies, or identifies, possibilities. (For an overview of these issues, see Dasgupta 2015; Hawthorne and Sider 2002.) The question is rather how to understand the truth-conditions of ordinary attributions of rest and motions. Even if there were some theoretically motivated notion of absolute rest, perhaps applying to the centre of mass of the universe, this would be irrelevant to the truth or falsity of ordinary judgments that something is moving or staying in the same place.

the two communities would soon reveal that their methods, though apparently conflicting, were in fact equivalent and in substantive agreement.

We may want to gloss this equivalence by saying that the two communities' different celestial mechanics state the same underlying facts, or have the same truth-conditions. But what are these facts or truth-conditions? There is some temptation to think that, in order to make sense of the idea of a domain of underlying facts, they should be statable in the terms of a theory that does not itself require relativisation to any frame of reference. In the case of attributions of rest and motion, we do indeed have a clear idea what this would be: a theory in which all statements of rest and motion are couched in terms of changing spatial relations between bodies, rather than changes in absolute spatial position. This can either be a fully relational theory, as was envisaged by Leibniz, or a 'Galilean' theory of spacetime, which makes an absolute distinction between constant and accelerated motion, but none between rest and constant motion.¹⁰

On the other hand, it is not obviously required for seeing two different indexical sentences used in different contexts as stating the same fact that one be able to state their common truth-conditions in a metalanguage itself free of indexical devices and interpretable uniformly in any context.¹¹ It may well be that reference to spatial objects is only possible with the use of devices that, at some level, exploit one's spatial location. This is implied by P. F. Strawson's (1959, pp. 20–23) argument from 'massive reduplication': if all qualitative features can be duplicated, there may in principle be no uniquely identifying description of an object or its location purely in terms of its features. If this is so, then the only way to uniquely single out an object or location is with the use of indexical devices that exploit one's own relations to the thing singled out. Thus there may be no canonical level of description of reality available at which all use of indexical devices falls away. But this should by no means lead us to conclude that indexical sentences used in different contexts cannot be truth-conditionally equivalent. This point will recur a number of times in subsequent discussion.

It is important to note that the notion of fact, and of truth-condition, under discus-

¹⁰For an introductory discussion of relational and Galilean spacetime, see Dasgupta (op. cit.), Maudlin 2012, ch. 2.

¹¹For the idea that a semantic theory for indexicals should use indexicals in the metalanguage, see Rumfitt 1993.

sion may well go beyond what is strictly part of the semantics of the relevant fragment of discourse. In giving a semantics for predicates of rest and motion one might simply adopt whichever spatial frame of reference is the standard one for speakers of the language, and give purely disquotational truth-conditions. This would be appropriate in particular if one wishes for a semantics to be *interpretive*, in that sense that its truth-conditions capture, at least roughly, speakers' own understanding of the meanings of their sentences.¹² Even if we can in principle give a purely relational description of the state of affairs described by the sentence 'Jupiter is in transit' as used at a particular terrestrial location, this may not be any part of what ordinary speakers understand by the phrase: what they understand is, rather, the apparent motion of a visible object across the night sky. Nevertheless, there remains a clear intuitive sense in which descriptions of celestial phenomena from different frames of reference are just different descriptions of the same thing.

What secures this sense is the underlying systematicity of our thinking about the spatial domain. Implicit in our understanding of spatial statements is the idea that the spatial framework constitutes a single arena in which all physical processes occur, and that the spatial arrangement of physical things remains constant through changes in an observer's relative spatial position. These features ensure that our spatial conceptions can survive adjustments that introduce relativisations to hidden parameters in order to preserve the unity of the framework. The fact that observers with different frames of reference make apparently conflicting observations is an apparent challenge to this sense of underlying unity; but the challenge can be comprehensively disarmed by adopting a relational framework in which the apparently different states of motion and rest of the same bodies recorded by different observers can be fully accounted for in terms of the relative motion of those bodies and the observers' frames of reference. This procedure makes good on the idea that the observers are occupying different perspectives on a single spatial reality.

Moving to a relational conception of rest and motion is not an utterly radical revision of our ordinary ways of thinking. As noted above, it is to some extent already

¹²This ambition is encapsulated by Donald Davidson's (1984d) conjecture that a theory of truth for a language can 'do duty as' a theory of meaning. Useful discussions of Davidson's project include Evans and McDowell 1976a, Larson and Segal 1995, ch. 2, Neale 2001, ch. 2.

implicit in those ways of thinking that something may be at rest when viewed from one standpoint and moving when viewed from another. But more major revisions of our spatiotemporal notions are possible; the discovery of Special Relativity that distal simultaneity is well-defined only relative to a frame of reference is, arguably, one such example. This does appear to be a genuine advance in objectivity of the kind that Nagel and Williams describe.¹³ But again, what makes this revision intelligible is the underlying unity and systematicity that was already present in our thinking about space.

The case of attributions of motion and rest thus seems to give us a paradigm of the kind of explanation which Williams envisages as revealing a theory or representation as depending on a local perspective. The two components here were, first, the possibility of multiple, apparently conflicting descriptions being given from different perspectives with different frames of reference; and, secondly, the availability at least in principle of a theory of the underlying structure of the domain that does not privilege any frame of reference—in this case, a relational or Galilean theory of spacetime.

It should be clear that the case of explaining the apparent colours of things in terms of underlying physical processes is nothing like this. First, we have nothing like the plurality of apparently conflicting perspectives we had in the spatial case. To the extent that there are differences among judgments of colour, it is typically open to construe these at face value, as genuine disagreements about the colours of things. Secondly, the role played by the ‘perspective’ of the human visual system in the explanation of colour is fundamentally different from the role played by spatial location in the explanation of apparent motion in relational terms. The latter amounts to a revision in our existing conception of the spatial domain, introducing an extra layer of generality to the interpretation of existing attributions of rest and motion. The relevant notion of perspective, in this case a spatial frame of reference, is already provided for within the revised conception. The former, on the other hand, involves replacing the domain of colours with something that is conceived of as entirely different, namely reflectance profiles and light wavelengths. Recovering the apparent facts about colours thus requires introducing a *new* kind of ‘perspective’—the constitution of the human visual system—that is extraneous to anything given by the underlying physical theory of light wavelengths.

¹³Cf. Nagel, p. 76

With regard to the first point, one might introduce a proxy for the existence of actual differences in judgment by considering hypothetical scenarios. Thus A. W. Moore writes,

Suppose...that there are aliens who are so physiologically constituted that a particular substance looks red to them though it looks green to all normal-sighted human beings: the substance affects their visual apparatus differently from ours, though they otherwise make the same discriminations as we do. Imagine now that one of these aliens classifies the substance as red while one of us classifies it as green. There need no more be a conflict than in the case where one person remarks, 'It is snowing', and another remarks 'It has stopped snowing', or indeed the case where an alien says that our sun is thousands of light-years away and one of us says that it is approximately 93 million miles away. Both classifications may be correct, the one because of its alien location, the other because of its human location. (A. W. Moore 1997, p. 5)

Moore hopes in this way to introduce an abstract notion of 'location' (or 'point of view') broadly analogous to the notion of a spatial location in the role it plays in accounting for the truth of ordinary statements. The relevant aspects of location in this case are, presumably, facts about the alien and human visual systems.

Contrary to what Moore says, it is far from obvious the judgment that there is no real conflict here would be forthcoming in such a case. We would probably be tempted to say that the aliens were subject to a quite specific illusion with respect to the colour of the substance in question. Of course, typically in cases of perceptual illusion we expect there to be some explanation of why it occurs. If no such explanation was revealed by extensive psychophysical research—if everything appeared to be functioning normally with regard to the aliens' processing of light wavelengths reflected from that specific substance—this might eventually undermine our confidence in our own perceptual judgments that the substance was green. We might come to suspect that really the substance was neither red nor green, or perhaps lacked a colour in the normal sense altogether. Even at this point, there is no temptation to say that both judgements are correct. Rather, we might suspect that both were false.

Of course, if *this* final attitude of scepticism were generalised to other judgments of colour, that would be a different way of making out the conclusion that things do not really have colours intrinsically. We might then speak as if they did for everyday

purposes—in which case we could, at a more theoretical level, say that ‘red’, as used by a human, is true of something just in case that object looks red to humans (and *mutatis mutandis* when used by one of the aliens.) This would perhaps amount to a genuine revision in our conception of the domain of colours, somewhat analogous to the move to a relational conception of space. But notice that the case as by described by Moore does not provide any grounds for that general sceptical move, nor any corresponding revision in our conception of the domain of colours. Rather, what we should say in Moore’s case is that we have a local aberration in the chromatic properties of a specific substance, such that it cannot be properly described either as green or red. The idea that objects which reflect light in unusual or variable ways cannot stably be ascribed any single colour is by no means outlandish. But it does not undermine ordinary attributions of colour to other, more chromatically well-behaved objects.

On the other hand, in a case where there was widespread and systematic (apparent) disagreement between two communities, not just about the colour of one specific object, but about the colours of things in general, there might be grounds for the generalised sceptical move. The problem is that it is not clear whether such a case is coherent. Identifying disagreement is possible only against the background of widespread agreement. So we can imagine two communities who make broadly the same judgements about colour, but in some specific cases diverge—this is Moore’s case. But it is not clear we can suppose widespread divergence in judgments, and yet hold onto the idea that there is even an apparent disagreement.

Suppose an alien community group objects differently from us with respect to colour: that is, they make different judgments from us about which things are similar or different in colour. Then we would think of them as using a different repertoire of colour concepts from ours, ones for which we do not have ordinary words. The concepts of our respective repertoires would overlap and cut across one another, and so we would apply different colour concepts to the same object. But here there is not an apparent disagreement to be resolved by supplying a hidden contextual parameter. Rather, we would just be talking past each other.¹⁴

The underlying phenomenon here is that the concepts which feature in ordinary

¹⁴This point is made clearly by Hyman 2005.

colour judgments are relatively course-grained, whereas colours themselves admit of continuous gradation. Whenever a continuous domain is described by means of discrete concepts, there is room for arbitrary differences in how to carve it up.¹⁵ There may be some explanatory story of why two different communities carve things up in a different way. With us and the aliens, such a story will likely advert to differences in sensory apparatus. We may want to call this a difference in perspective. But ‘perspective’ in this sense just means those features of a thinker’s situation which explain why their concepts track certain differences in nature rather than others. It does not deliver any sense in which the *correctness* of a representation which uses those concepts depend on perspective.¹⁶

Dwelling on these issues brings out just how tenuous the analogy is between a spatial perspective and the kind of sensory ‘perspective’ on which ordinary colour discourse is proposed to depend. The fundamental point is that there is no element of perspective built into ordinary thinking about colour the way there is for space. It is already implicit in even the most rudimentary spatial thought that the same spatial array can be represented differently when viewed from different locations. There is just no analogue of this in the case of colour. Rather, experience purports to reveal to us the colours of things, full stop. This is not to deny that there is complexity in colour experience, that the same object can appear differently under different lighting conditions, and so on.¹⁷ But such

¹⁵This should be not confused with another way in which two speakers or communities may differ, namely in respect of the standards of precision applied when using inexact concepts. This, by contrast, does introduce a genuine need to relativise to a contextual parameter. But it is not what is going on when two community or species make different colour discriminations.

¹⁶Sometimes there may be room for higher-order disagreement about what is the correct way to carve things up. Often this may be taken at face value, as a genuine disagreement over the organisation of nature. For instance, the idea of natural kinds as forming trees or hierarchies provides for genuine disagreement over how concepts in the lower levels of the hierarchy should be grouped. There is nothing mysterious about saying that the classification of whales as fish, or of humans into the racial categories that emerged around the 19th century, are just incorrect. (For the latter claim, see Appiah 1994, pt. 1.) By contrast, it is far less plausible to insist on any uniquely correct way of grouping the domain of colours. A more controversial question is whether the classification of the natural world reflects, in some deeper way, local practical interests. This would introduce a genuine relativity to interests, since what is in question is the higher-order correctness of the groupings themselves, rather than the first-order correctness of representations involving concepts that express those groupings. Addressing this question would take me too far afield. For now I just want to insist that it must not be confused with the idea that the first-order correctness of, for instance, colour statements, is relative to a point of view.

¹⁷For discussion of colour constancy phenomena, see Hilbert 2005.

complexity as there is does not necessarily support the idea of a hidden parameter on which the truth of ordinary monadic ascriptions of colour might be seen to depend. The conception of colour we obtain from experience is not one which allows for the idea of some underlying phenomenon which may be viewed differently from different sensory ‘perspectives’.

Moore concedes something very like this point in his own discussion. But he does not take it to undermine the idea that colour discourse is nevertheless in some sense perspectival. He writes,

Even so, the argument serves to remind us of how our colour concepts are responsive, at some fundamental level, to the look of things. Ordinary uses of ‘green’ may not involve a suppressed relativization (‘green for humans.’) But beings whose visual apparatus differed radically from ours, or who had no visual apparatus at all, would be unable to grasp our colour concepts. Even if they could measure wavelengths, and even if they could tell, indirectly, that something was green, they would still not understand this in the way we do. They might well not be able to see the point of our classifications. To that extent our use of colour concepts is dependent on its location. There is still a sense in which, when I say, ‘Grass is green,’ I am producing a correct account of how things are from a human perspective. (Ibid., p. 6)

But the point that wielding colour concepts requires having a certain sensory apparatus is not by itself sufficient to establish that colour discourse is perspectival in the sense that interests Moore. It is no less true of the concepts used in mathematical physics that, in order to possess them, one has to satisfy certain enabling conditions. Moore’s idea must be that the sensory conditions of grasping colour concepts are somehow more idiosyncratic or parochial. One way of establishing this, discussed in the previous section, would be to give a satisfying explanation of the perspective from which things appear coloured, and of why they appear coloured, in purely physicalistic terms, without the use of colour concepts. This is what Stroud (op. cit.) calls an ‘unmasking explanation’ of colour experience.

Thus even if the analogy with spatial perspective is strained, there may yet be some meaningful sense in which colour concepts are tied to an idiosyncratically human perspective. The point is that we will not arrive at this conclusion just by reflecting on the structure and correctness conditions of colour discourse. In the spatial case an idea of

perspective was already built into our original understanding of the domain. By contrast, any sense in which colour concepts are perspectival can only be achieved through adopting a more 'sideways-on' stance, one which abstracts from the familiar ways of thinking to which the conception of things as coloured belongs, and offers an explanation of the subject-matter of colour discourse in colourless, wholly scientific terms.

1.3 THE EXPLANATORY GAP AND THE METAPHYSICS OF PERSPECTIVE

There is a basic problem which confronts anyone who attempts this task. An unmasking explanation of colour discourse necessarily goes beyond the underlying physical theory of light wavelengths, and seeks further to explain why things appear coloured to human observers. Explaining these psychological facts is the whole basis of the claim of the scientific conception to transcend, and thus reveal as idiosyncratic, the everyday conception of things as coloured. The problem is that it is very difficult to see how such an explanation could go without ultimately using the very colour concepts it seeks to supplant. Stroud puts the difficulty succinctly:

It is true that the physical-scientific story is enough to explain Jones' physical interaction with the lemon. But what is to be explained in a case like this is the fact that Jones sees yellow and sees the lemon to be yellow. Since those psychological facts are not equivalent to anything expressed in purely physical-scientific terms, an austere scientific explanation of the state of Jones' brain in the presence of the lemon does not explain why she sees yellow there. It explains only something expressible in purely physical-scientific terms. This is a consequence of the assumption that there are no psychological and no colour terms in the vocabulary of the physical sciences, combined with the idea that explanations explain facts, not objects. (Stroud 2000, p. 177)

The unmasking explanation assumes it is built into the idea of a physical science that sensible qualities such as colours are excluded from the world it describes. The idea is then that the experience of colour could nevertheless be explained in scientifically acceptable terms, without employing the notion of colour itself, for instance as a neurophysiological or informational state. The challenge raised by Stroud is that, even if such a physical science can explain the occurrence of certain states, it has not thereby explained the experience of colour. In order to secure the idea that what is explained is

indeed an experience of something as yellow, we need to identify it as just that in the explanation—as an experience in which an object is seen to manifest a sensible quality of a kind foreign to the physical sciences. There is an explanatory gap between the world as described by the physical sciences, and the world as it shows up in experience.¹⁸

This kind of explanatory gap leads Thomas Nagel to the conclusion that any objective conception of reality is necessarily incomplete. In *The View from Nowhere* and other work, Nagel repeatedly returns to the idea that there are features of reality which are inextricably tied to a specific point of view. He reaches this conclusion for a number of domains—free will, value, morality—but he makes the case most forcefully for subjectivity and consciousness. On the model of objectivity advanced sketched earlier, we come to understand a phenomenon better by abstracting away from the particular perspective we have on it. But, he argues, things are different when it comes to consciousness:

Experience itself, however, does not seem to fit the pattern. The idea of moving from appearance to reality seems to make no sense here. What is the analogue in this case to pursuing a more objective understanding of the same phenomena by abandoning the initial subjective viewpoint toward them in favour of another that is more objective but concerns the same thing?...If the subjective character of experience is fully comprehensible only from one point of view, then any shift to greater objectivity—that is, less attachment to a specific viewpoint—does not take us nearer to the real nature of the phenomenon: it takes us farther away from it. ('What Is It Like to Be a Bat?', p. 174)

Nagel's key thought here is the conscious character of experience is essentially connected with its being from the point of view of an experiencing subject. But, it is just the operation of abstracting away from a subject's point of view that is characteristic of the kind of objectivity exemplified by the natural sciences. As such, any objective account of experience will necessarily leave behind the point of view which was partly constitutive of its nature. From this Nagel reaches the quasi-paradoxical conclusion that consciousness is something of which it is impossible to reach a fully objective understanding.

The recalcitrance of experience and consciousness to the kind of objective understanding exemplified by the sciences leads Nagel to argue for the reality of irreducibly perspectival facts:

¹⁸This is of course related to Jackson 1982's famous thought experiment involving Mary the colour scientist.

We must admit that the move toward objectivity reveals what things are like in themselves as opposed to how they appear...Therefore when the objective gaze is turned on human beings and other experiencing creatures, who are undeniably parts of the world, it can reveal only what they are like in themselves. And if the way things are for these subjects is not part of the way things are in themselves, an objective account, whatever it shows, will omit something. So reality is not just objective reality, and the pursuit of objectivity is not an equally effective method of reaching the truth about everything. ('Subjective and Objective', pp. 212–213; cf. *The View From Nowhere*, ch. 1; pp. 25–27.)

It is important to stress that Nagel is not just saying that the facts about consciousness are facts which concern the point of view of a conscious subject. Rather the idea is that there are facts which, by their very nature, can only be grasped from a certain point of view—facts of how things are 'for a subject', rather than in themselves. This is not meant to be a merely epistemic restriction, for instance on possession of the requisite concepts to grasp those facts. Rather, we might say that such facts obtain only within a point of view, or belong to that point of view. This is, as it were, a metaphysics of points of view.¹⁹ Facts about colours and other sensible qualities are a paradigm example of the kind of essentially perspectival fact Nagel has in mind: facts which obtain only within, or from, the point of view of a creature with a certain sensory apparatus.

Moore argues that Nagel's notion of a perspectival fact is incoherent. His argument is basically this: If there are perspectival facts, then there can be conflicting facts belonging to different perspectives. For instance—to revert to the earlier example—the book is on the left from my perspective, and on the right from yours. This can lead to the idea that you and I inhabit different worlds, that there is no single reality containing both of our perspectives. If, on the other hand, we want to retain the sense of a common reality, we have to say something else: for instance that the book is on *my* left and *your* right. But this removes the sense in which the facts are genuinely perspectival: what we have is instead different perspectival representations of the same fact. Moore puts it like this:

...the pressure to say that there are different worlds (and thereby deny the unity of reality) arises from a hazy sense that one world cannot accommodate all the

¹⁹For some more explicitly theoretical statements of a Nagelian metaphysics of perspectival facts, see Hare 2007; Lipman 2016; Merlo 2016. A specific form of perspectival metaphysics, which I will discuss in ch. 2, is realism about tense. The authors just cited tend to take tense realism as in some sense the prototypical case.

different perspectival facts that are supposed to obtain. It cannot. Unless one acknowledges that the obtaining of these facts is relative to a point of view, one always, ultimately, lapses into incoherence...On the other hand [Nagel's view] requires that the obtaining of perspectival facts should *not* be relative to a point of view. It requires that perspectival facts should simply obtain, and, as with other facts, that their obtaining should consist in the world's being a certain way. (Op. cit., p. 46)

Moore concludes that being perspectival, or from a point of view, is a property of representations, not facts. But what are the nonperspectival facts that are perspectivally represented, say, by creatures with colour experience? Moore thinks they are physical facts, for instance about the microphysical properties of reflective surfaces:

...in [the case where Alison sees a green apple] the apparently colourless scientific story *is* the whole story. The apple is green, certainly. Alison correctly represents it as such. But I am able to say that because I am able to represent it from the same sensory point of view. I have indirectly endorsed Alison's representation. But I could have done so just as well by telling the scientific story. There is nothing that is only there from her sensory point of view. The fact that the apple is green is not a perspectival fact. (Ibid., p. 45)

But now the question arises again just on what basis the fact stated by Alison when she says the apple is green is identified with some physical fact about its reflectance profile, as long as this fact is assumed not itself to be a fact about colour. Moore wants to say that saying the apple is green is just a way of representing a non-chromatic, physical fact from a distinctively human perspective. But the point pressed in the above section was that our conception of colour does not make available to us a notion of a sensory perspective on physical reality anything remotely like the idea of spatial perspective implicit in ordinary thinking about the spatial realm.

Moore's story about the colour of the apple thus brings into play a properly metaphysical notion of a fact. The idea that the fact which we express by saying that something is green could be identified with a colourless, purely physical fact, takes us beyond any notion of facts and their identities that can be extracted from the truth-conditions of colour discourse, at least insofar as the specification of those truth-conditions is constrained by speakers' own understanding of the domain of colours. Rather, it must be built on some prior commitment to the idea that the underlying facts which explain the

truth of our discourse are colourless ones. But at this stage of the dialectic, this looks like a bad piece of dogmatic metaphysics. Why should we think that colours do not feature in the facts of reality, absent some positive reason for excluding them?

Moore would presumably say: because the only alternative is countenancing Nagelian perspectival facts, with all the difficulties that entails. We thus seem to be presented with a dilemma: either allow that reality is irreducibly perspectival, in which case we are confronted with its fragmentation into different incommensurable ‘worlds’, one for each point of view; or else follow Moore in holding that perspectival representations are representations of a nonperspectival reality, and simply tolerate the explanatory gap between the facts in themselves and our perspectival representation of them.

In fact, however, the putative crisis is overstated. The culprit is once again the obscurity inherent in the notion of representation’s being from a point of view, when taken beyond the paradigmatic spatial case. It is true in some sense that facts about the colours of things are proper to the point of view of a creature constituted a certain way. But this is not necessarily to say any more than that facts about the colours of things can only be apprehended by creatures who can perceive colour.²⁰ To be able to perceive colour, certain things have to be true of a creature; in particular, they have to be equipped with the right sensory apparatus. But it is true of more or less any fact that in order to possess the concepts required to grasp it, certain things have to be true of one. The sense in which colour concepts depend upon perspective may not be one which implies that they fail to represent reality.

1.4 ABSTRACTION AND EMBEDDING

As was noted several pages back, there is a perfectly legitimate sense of perspective in which one’s perspective determines what features one is aware of and in what manner. This notion of perspective is, in P. F. Strawson’s phrase, a creature’s ‘experiential route’ through the world it inhabits.²¹ A core aspect of a creature’s experiential route through

²⁰Even this may be an exaggeration. Cf. Stroud, *op. cit.*, p. 173

²¹Strawson 2006, p. 104: ‘But it is a shining fact about such a series of experiences...that its members collectively build up or yield...a picture of a unified objective world through which the experiences themselves constitute a single, subjective, experiential route, one among other possible subjective routes through the same objective world...The history of a man, we might say, is—among much else—an em-

the world is their literal spatial position. One's spatial position determines, both, what one can actually perceive, and how those things appear to one. More broadly we might include in a creature's perspective the background features of their perceptual environment, such as ambient noise, temperature, lighting conditions, and so on, as well as the perceiver's sensory discriminatory capacities; and perhaps also such factors as background beliefs, education, cultural upbringing, and so on.

It should be clear, though, that the dependence of a subject's representations on their experiential route through the world need not entail that those representations are perspectival in the sense that their correctness depends, in part, upon unrepresented features of the subject's perspective, such as their spatial location and orientation, as well as on what is represented. This plausibly is the case, at least, for some aspects of the spatial content of visual experience, for instance representations of things as being far away, or on the left. This kind of perspective-dependence is mirrored in the semantic features of indexical devices. When a representation is perspectival in this sense, we might say it is *egocentrically perspectival*.

But there is an equally good sense in which the dependence of a representation on a creature's perspective amounts just to the fact that its occurrence was a consequence of how and where the creature was situated vis-à-vis the represented subject-matter. This is no more than the observation that to perceive something, one has to be in the right place at the right time. More generally, what a creature is aware of depends on what sensory capacities it has. I will say that representations which depend on a thinker's perspective are *positionally perspectival*. But there is no attendant sense in which the *content* of a representation thereby depends on, or is relative to, its subject's position in the world, or its sensory capacities. Rather, these features of the creature's perspective are just enabling conditions of the representation. We may contrast both of the above kinds of perspectival representation with Nagel's notion of an essentially perspectival fact. Representations which are of such Nagelian facts we might call *radically perspectival*.

This notion of perspective as an experiential route, or the determinants of an ex-

bodiment of a temporally extended *point of view* on the world.² Note the difference between the notion of a point of view as Strawson articulates it here and Nagel's. The sense in which Strawson associates consciousness and experience with having a point of view, in the sense of an experiential route, is that experience must provide for the possibility of multiple points of view on the same. Whereas Nagel's idea that consciousness is necessarily tied to a specific point of view makes it, in an important sense, *private*.

periential route, is surely key to the sense in which consciousness and subjectivity are perspectival. Naomi Eilan puts it like this:

Such world-dependence [i.e. of a subject's experiential route through the world] constitutes a kind of subjectivity, contrasted with the objectivity of representations from no point of view. Such [subjectivity]²² is an essential feature of what makes a series of representations constitute a conscious experiential perspective on the world. Experiential consciousness is subjective, in part, precisely in the sense that the contents of a conscious perspective are, so to speak, glued to the ground, and in this sense immersed, in contrast to the lofty detachment of absolute conceptions. (1997, p. 243)

Acknowledging this sense in which consciousness is perspectival is a way of disarming the dilemma which emerged between Moore and Nagel. They share the assumption that certain facts, such as facts about colours, are, as Moore puts it, 'responsive, at a fundamental level, to the look of things', and both take this to imply that they cannot be simply be facts about how things objectively are. This common assumption leads Nagel to posit a realm of autonomous, radically perspective facts; on the other hand the physicalist reduction of sensible qualities beckoned because it seemed the only way in which the facts as they appeared to a conscious subject could be integrated into a wider reality where the point of view of the subject figures as one among others. But if we instead characterise the perspective of a conscious subject as an experiential route through the world, in terms of the objects and qualities of which the subject is aware, then we no longer need to think of consciousness as a kind of perspectival filter or veil on an independently characterisable reality in order to anchor the subject's point of view to the world. We no longer need to cast about to articulate a notion of a sensory perspective on an underlying physical reality, conceived by means of an overstrained analogy with the familiar notion of a spatial point of view.

We can correspondingly distinguish two very different projects whereby subjects might seek to make sense of their relation to the world, embodying two different conceptions of an objective advance. Christoph Hoerl (op. cit., pp. 198) helpfully labels these the 'abstraction conception' and the 'embedding conception'. On the abstraction conception, one comes to realise that what one took to be features of reality are in some

²²Original text reads 'Such objectivity...' The context leads me to suspect that this is a typographical error—although I am ready to be corrected on this.

sense an artefact of one's manner of representing it. This characteristically involves an understanding of the content of one's representation as being made true both by its subject-matter and by features of one's perspective. For instance, the perception of the sun's motion across the sky is seen to be correct in virtue of, first, the revolution of the earth around the sun and, secondly, the location of the perceiver on the surface of the earth. It is not just that the perceiver's location explains how they were able to see the sun's motion. Rather, their location provides an explanation of how their perception can be correct, even though the sun does not *really* move across the sky. The sort of explanation of colour discourse I have been considering attempts to generalise these strategy to explain the appearance of colour as the joint upshot of non-chromatic properties and features of the human visual system. But, as I have been pressing, it is unclear that this is a tenable conception of the subject-matter of colour experience.

By contrast, the embedding conception involves seeing a subject's perspective as merely providing enabling conditions of their being in a certain kind of state. As Hoerl puts it, the idea is that 'we achieve an objective understanding of an individual's point of view, one that does not depend on occupying that point of view oneself, by looking at how that individual is embedded in the world, at the particular relations in which it stands to features of that world, and by explaining particular aspects of its point of view in terms of it standing in those relations to those features.' (2015b, p. 195)²³ Applying this embedding strategy to one's own states and experiences, in contrast to the abstraction conception, does not involve any revision in one's theories of the kinds of features and entities that the world contains. It does not purport to undermine one's representations' claim to reality.

This is not to say that the embedding conception is metaphysically neutral. On the contrary, it brazenly appeals to the sensible qualities themselves in explaining the perspective of a conscious subject who is aware of those sensible qualities, and hence offers no prospect of making good on the suspicion that such sensible qualities are not part of of reality as it is in itself. In the case of colour, the commitment is clear: it requires countenancing colours as intrinsic, non-dispositional properties of objects.²⁴ This does not rule out that such properties might be non-fundamental, in some sense yet to be made

²³See also Campbell 1993; Eilan 1994, 2014 for related implementations of this strategy.

²⁴Campbell (op. cit.) is a classic statement of this position.

out; or that some basis might be produced for identifying them with physical facts. The point is that any such basis will not come from an account of how things' appearing coloured depends upon subjectivity.

A corollary is that the embedding strategy may come into conflict with prior metaphysical commitments that block countenancing the reality of colours and other sensible qualities. A prior commitment to austere physicalism—perhaps based in considerations of the unity of science—might generate such a conflict. A different source of conflict—one that will become relevant in the discussion of tense and temporal perspective—would be if there is no consistent way of describing the world as containing the materials required by the embedding strategy. Employing the embedding strategy in cases of inverted spectra, for instance, might be argued to commit one to the possibility of an object being both red and green. This argument is of course not decisive, and it is open to the proponent of the embedding strategy simply to deny the possibility of inverted spectra, understood as the argument requires.²⁵ The current point is not about that specific argument, though, but rather that the successful employment of the embedding strategy does seem to require a certain kind of consistency in the worldly materials it appeals to.

I began this chapter by noting a highly general metaphysical question, or anxiety, that arises through reflection on the fact that thought and experience depends, in some sense, on our idiosyncratic perspective or route through the world. The main aim of the discussion has been to distinguish different ways in which a subject's representations may be said so to depend.

In the first place, what I have called egocentrically perspectival representations are distinctive in that their content, is the joint upshot of what is represented along with local features of the subject's perspective of which they may not be consciously aware. In this case, explaining why a representation is true or correct will involve bringing out or making explicit a hidden contextual parameter. The paradigm of this is the kind of perspective involved in egocentric spatial representations, such as in terms of left and right. Recognising previously unacknowledged contextual parameters typically means

²⁵This is the strategy of Campbell (*op. cit.*)

a revision in a subject's conception of the represented domain, as when one acknowledges the relativity of rest and motion to a frame of reference. The general project of revealing hidden indexicality in this way I called, following Hoerl 2015b, the abstraction conception of objectivity. This conception seeks to improve a subject's theory of the world by revealing, or 'unmasking', certain features of it as mere artefacts of perspective. When the abstraction conception fails to have some suitable story available to explain the appearances we may be feel the pull of Nagel's idea of a realm of facts constitutively tied to a certain perspective.

On the other hand, a representation may be merely positionally perspectival, in that the subject's having of it is enabled by their being in a certain location, or perhaps in their enjoying certain sensory capacities. Making sense of how one's representations depend on their experiential route through the world in this sense is what I called the embedding conception of objectivity. Unlike the abstraction conception, the embedding conception does not seek to undermine the reality of appearance. Rather, it involves a commitment to the world containing, by and large, the features and things it appears to contain.

A especially rich and interesting version of the embedding strategy, which I will be exploring in the coming chapters, involves the idea that the perspective of a conscious subject involves, in a fundamental way, an awareness of particulars. In characterising a subject's experiential route through the world, we typically mention not just the repeatable features and qualities which contribute to their experience, but the particular things which instantiate those qualities. On this version of the embedding strategy, reference to such particulars is not an accidental feature, but an essential ingredient in how to make sense of the subject's relation to the world.

This role for particularity in constituting a subject's perspective moreover constitutes an important point of contact between thought and experience. Beliefs concerning particulars are tied much more closely to one's experiential route through the world than are more theoretical beliefs. The paradigm case is that of a demonstrative thought about an object of which one is presently perceptually aware. Of course there are other, less direct ways of thinking about particulars, for instance by exercising one's competence with a proper name. But nevertheless singular thought is anchored to experience—'glued to

the ground', in Eilan's phrase—in a far more immediate way than general theories and representations.

I noted before that the general metaphysical question of the reality of appearances may be posed at the level of discourse or theory, asking whether certain concepts or bodies of belief answer to anything in reality. In posing the question in this way, we are encouraged to fix on general beliefs, such as 'grass is green', 'murder is wrong', and so on, in order to ask how they correspond to reality. This approach is especially friendly to the abstraction conception of objectivity, in which the goal is to improve a subject's overall theory of the world by removing from it those elements which are merely the contribution of their local perspective. It can then seem as if the only role experience can play is that of a filter or veil, making things seem otherwise than how they really are. By contrast if we understand our manifest image of the world to include, in addition to theories and generalisations, a sensitivity to particularity, we will be able to see how a subject's contingent and idiosyncratic relations to their environment contribute to their take on the world without thereby seeking undermining the claim to reality of their specific beliefs and experiences.

With these points in mind, I will turn in the next chapter to considering how this same dialectic of perspective and reality plays out in our understanding and experience of the temporal world.

2 | TENSE AND TEMPORAL PERSPECTIVE

The previous chapter sketched a general dialectic that arises on asking the metaphysical question of how, or whether, a certain area of thought or discourse corresponds to reality. The sense that a certain concept or distinction fails to pick out anything in reality was closely tied to the idea that its use is tied to an idiosyncratic or local perspective, and hence that an absolute, perspective-free representation of reality would not operate with such concepts. However this line of thought was argued to involve an equivocation over the notion of perspective in play. If a representation is egocentrically perspectival—meaning that an explanation of its truth or correctness necessarily includes features of the subject’s perspective, as well as those of the world thereby represented—then there may be a genuine sense in which at least some of the concepts employed in that representation fail faithfully to reflect reality. If, on the other hand, a representation is positionally perspectival—meaning merely that a thinker has to be in a certain position, or meet certain conditions, in order to be grasp that representation—then no such conclusion is motivated.

This chapter will explore how this dialectic applies with regard to our experience and knowledge of the temporal world. A persistent theme in philosophical discussions of time is that there are features of our naïve theory of time—‘manifest time’, as Callender (2017) labels it—whose metaphysical status as reflective of reality is questionable. The aspect of manifest time with which I shall be principally concerned is the idea that the past is fixed and the future is open. Explicitly articulating our naïve conception, we appear to be committed to the following two thoughts:

Difference: There is a categorical difference between past and future events: the past is fixed, and the future is open.

Variation: What is past and what is future changes.

There is at the very least a *prima facie* tension between Difference and Variation. Difference states that the past-future distinction tracks a categorical difference in the status of events either side of the division. This would seem to require that the difference be absolute—rather than relative to a time. On the other hand, the only apparent way to make sense of the change required by Variation is that what is past and what is future changes according to which time is present; and this would seem to require taking the relevant difference to be one that obtains only relative to a time. We seem to require the past-future distinction to be both relative and non-relative.

This is one version of a general difficulty that characterises reflection about the metaphysical status of tensed representations. Our ordinary conception of time seems to require both that distinctions of past, present and future be absolute, and that they apply differently at different times. The understanding of time expressed through the use of the tenses thus makes conflicting demands on our conception of the reality thereby represented.¹ This gives a quite distinctive shape to the metaphysical question as applied to the temporal world.

These matters interact with another feature of manifest time, namely the idea that time passes. The passage of time has often been thought to make particular demands of temporal reality, so that that one must either adopt a particular metaphysics of time, or declare the passage of time to be an illusion.

One aspect of this question I shall largely set aside, namely that a metaphysics of passage is required by the fact that temporal experience has a ‘flowing’ or ‘animated’ character. One reason to do so is that this feature of temporal experience brings in further issues about physical continuity which tend to muddy the waters. It is not clear, at least to me, that in pointing to the flowing character of temporal experience we are identifying any feature further to the fact that we experience change as continuous rather than discrete.² Issues of time and the continuum are deep and subtle, but their connection

¹This tension is at the heart of McTaggart’s argument for the unreality of time (Dummett 1978a; Fine 2005; McTaggart 1900, 1908).

²For instance, L. A. Paul’s (2010) discussion of temporal experience aligns the distinction between experience of things in time as ‘static’ or ‘flowing’ with the difference between A- and B-theoretic accounts of change. But one might suspect that the distinction within temporal experience she identifies is

to questions about the reality-reflecting status of tensed representations is not obvious.

On the other hand, one might take the dynamical character of time to be a facet of a more general feature of temporality, namely that time is a dimension of genuine change, rather than mere variation. One way of articulating this difference is to insist that change, as opposed to mere variation, involves real contradiction between successive properties or states of affairs.³ One might then take the further step of understanding ‘robust passage’ to be a special and highly general kind of change, involving some distinctive such combination of contradictories—either time itself instantiating a series of incompatible properties, or temporal things changing in an ‘extraordinary way’, by instantiating a series of special, incompatible, temporal properties.⁴ This way of understanding the significance of dynamical character of passage does not particularly trade on intuitions that time is experienced as flowing; the thought that change requires contradiction would be just as compelling if all change were discrete.

Understanding the passage of time to require contradiction provides a route into questions of the reality of tense that does not necessarily go via the idea that there is an intrinsic difference between the past and future with respect to fixity. It is true that some philosophical theories of temporal passage—most notably, ‘growing block’ theories—understand the relevant change in terms that do directly relate to the asymmetry of fixity, such as the coming into being of new events, or the passage from mere possibility to actuality.⁵ But this is not required; it is at least open to someone to hold that the passage requires events to be successively future, present, and past, without attaching

really that between discrete and continuous change—a distinction which cuts across A- and B-theoretic accounts. This charge is pressed by, among others, Deng 2013 and Hoerl 2014.

³Cf. Dummett 1978a, p. 352: ‘Time involves change, and if there is change, then...some objects must have different predicates applying to them at different times...But this just means that to one and the same object incompatible predicates apply; for example, the paper was white and is yellow, so the incompatible predicates ‘white’ and ‘yellow’ apply to the paper.’

⁴The term ‘robust passage’, and this suggestion of two ways to understand it, is from Skow 2015, ch. 3.

⁵The idea that the passage of time consists in new events coming into existence is most closely associated with C. D. Broad (1923). More recent advocates include Button 2006; Correia and Rosenkranz 2018; Earman 2008; Tooley 1997. The alternative idea that past and future correspond to a difference in modality is a distinguished one, with roots in Aristotle’s logic and metaphysics. (Some expositions are Broadie 1982; Hintikka 1962.) The modern descendant of this line of thought is ‘branching time’ semantics such as Belnap and Green 1994; MacFarlane 2003; Prior 1967; Thomason 1970, which may or may not interpret the branching framework as providing a ‘robust’ notion of temporal passage. One example of an explicit attempt to explain the passage of time in these terms is the ‘shrinking tree’ theory of McCall 1976, 1994.

any particular weight to the idea that the future is open and the past is fixed.

Nevertheless, there is a more general interplay between issues about passage and the status of the difference between past and future. This is because the question of the reality of tense is fundamentally informed by intuitions concerning what kinds of demands the passage of time makes on temporal reality. Insofar as Difference does raise the question of the reality of tense, its assessment may partially turn on prior convictions about what passage involves. I will attempt to bring this out below.

My overriding goal is to investigate the notion of temporal perspective insofar as it coincides with that of an agent's perspective. My focus is therefore on the apparent difference in fixity between the past and the future, rather than the passage of time as such, since this aspect of our thinking about time plausibly has an intimate connection with our agency. One of my questions is whether making sense of an agent's perspective requires commitment to a tensed metaphysics. But I want ultimately to remain neutral about the question whether there may be other sources of commitment to such a metaphysics, and whether it can be made coherent.

With these points in mind, the rest of the chapter will review the main points of interest in the question of the reality of tense. Section 5.1 will discuss this in terms of the truth-conditions of tensed representations. There are systematic links between the truth-conditions of tensed representations tokened at different times, and so the question is how to respect this while insisting on the reality of tensed facts. Section 5.2 considers a more unashamedly metaphysical conception of the reality of tense, focusing on Kit Fine's recent reconstruction of McTaggart's argument. Section 5.3 moves the discussion away from metaphysics and back to the structure of temporal thought. I argue that the conflict between Difference and Variation should be understood as a tension within the perspective of ordinary temporal thought, rather than a pair of demands on a metaphysical conception of reality. This will prepare the ground for the more detailed discussion of the temporal perspective of agency in the coming chapters.

2.1 TENSE AND TRUTH-CONDITIONS

The question of the reality of tense is, To what extent do distinctions of past, present and future reflect a difference in reality?

As noted in the previous chapter, this question is elusive. Tense is, first and foremost, a feature of representations; and there is a real difference a tensed representation's being true or false—what more could we be asking beyond whether or not the tensed representations we take to be true, are in fact true?

We can understand this question as asking whether, and in what sense, tense is perspectival. The feature of tense expressed by Variation imply that tense must be perspectival just insofar as which events are past, and which are future, depends upon which temporal perspective one occupies. The most straightforward way of making sense of this involvement of perspective is to think of the tenses as indexically perspectival, in a manner broadly analogous to egocentric spatial notions such as left and right. Just as the same spatial arrangement can be accurately represented by differently located thinkers using different egocentric concepts, the same bit of temporal reality—the same series of events, say—can be accurately described by different tensed representations produced at different times. We have, in general, a pretty clear idea how this works: given some contextually fixed time of reference (in the case of verbal utterances, typically the time at which the utterance is produced), a past-tensed representation is true if the events it describes are located prior to the time of reference, and a future-tensed representation is true if the events it describes are located later than that time. (Cognate principles hold for metrical tensed concepts, like those expressed by 'tomorrow', 'an hour ago', and so on.) Understanding tensed representations to work this way, the only worldly structure we appeal to in accounting for their truth is a domain of events or times ordered by earlier-later (a B-series, in McTaggart's terminology.) And this brings us towards a picture of temporal reality where it is hard to make sense of Difference. We have left behind the idea that past and future constitute a difference in reality.

It bears remarking that the situation here is roughly the inverse of how things were in the case of colour. There we had a conviction, or suspicion, that distinctions of col-

our failed to correspond to any difference in reality; but nothing like the plurality of perspectives required to make good on an account of the truth of colour statements as depending on facts about the thinker's perspective. By contrast, in accounting for the truth of tensed thoughts, we are immediately confronted by a plurality of temporal perspectives on which the truth of tensed representations can be seen to depend—different tensed representations tokened at different times—but nevertheless remain convinced that the distinction of past and future, as made from each perspective, must pick out a real difference. But given the perspectival nature of tense, how can we even begin to make sense of this intuition?

There is little room for doubt about the need for some systematic story about how the truth of tensed representations depends on the time of their production. A semantics for temporal discourse should be able to give truth-conditions for representations produced at different times. This means recognising, first, that the same tensed representation-type ('It is raining') may receive different truth-values when tokened at different times; and, secondly, that there are systematic links between the truth-values of pairs of different tensed representation-types tokened at different times ('It is raining' as tokened on Monday, and 'It rained yesterday' as tokened on Tuesday).⁶ Acknowledging these basic facts seems to require us to think of the time of utterance as providing a contextual parameter that contributes to the determination of (tenseless) truth-conditions.⁷

Still, it might be felt that the sense in which there is a plurality of temporal perspectives does not yet provide for the idea that tensed representations are just perspectival ways of describing an underlying tenseless reality. The most that has been established is that tensed representation-types vary in truth-value over time, and that there are systematic links between the truth-values of tensed representations as tokened at different times. A semantics of tensed discourse typically accounts for these features by taking the time of each token representation to provide a contextual parameter that contributes to fixing the truth-conditions of that token. But this, it might be argued, does not establish that tense is perspectival in anything like the way that egocentric spatial rep-

⁶Cf. Dummett 1978b on tense and the 'truth-value links'.

⁷In fact evidence from natural language has been argued to support something stronger than this: it may be that tensed expressed need to be construed as referring to particular times or events, rather than merely determining truth-conditions jointly with information about the particular time of evaluation. Cf. Ogihara 2007; Partee 1973.

representations are. In particular, we do not necessarily have the idea that the same ‘bit of reality’—a tenseless fact—can be represented from different perspectives the way we do in the spatial case. Hence one may doubt whether the sentence ‘It is raining in London,’ produced on January 1, 2019, says the same thing, in the relevant sense of ‘same thing,’ as the sentence ‘There is rain in London on January 1, 2019.’

It is important to emphasise that taking tensed representations to have tenseless truth-conditions need not commit one to the thought that tensed sentences are cognitively equivalent to tenseless ones, or can be translated into them without loss of significance. That they are not so translatable can be seen from the fact that if someone wants to know the time, no tenseless sentence will suffice to inform them. This failure of translatability however is not peculiar to tense, but is a general feature of indexical representations: it is no less true of egocentric spatial representations in terms of ‘here,’ ‘left’ and ‘right.’ The fact that a tensed sentence is not cognitively equivalent to any tenseless one is consistent with the fact that what it says about the world, the way the world must be in order for it to be true, is a condition whose obtaining is tenseless, that is, does not vary over time. This divergence between truth-conditions and cognitive significance might motivate one to posit two ‘dimensions’ to the semantic value of an indexical expression and the sentences it occurs in: one encoding its links to perception and action, and one capturing its contribution to what is said about the world.⁸

Making the point this way, though, is liable to invite the suspicion that the tenseless so-called ‘truth-conditions’ are not really truth-conditions at all, at least insofar as the notion of a truth-condition is constrained by speakers’ own understanding of what is required for the truth of their statements.⁹ If we stick just to what I understand in using the sentence ‘It is sunny in London,’ perhaps all that is required for its truth is just that *it be sunny in London*. The relativisation to a tenselessly specified temporal parameter—‘on August 7, 2019’—can seem like a theorist’s imposition, required to preserve coherence between my statement now and statements produced at other times, but not in any deeper sense illuminating of what is thereby said about the world. This would be a piece of technical machinery, one whose introduction need not license the endorsement of a picture of tensed sentences as made true by an underlying tenseless

⁸E.g. Kaplan 1989; Lewis 1983; Perry 1979.

⁹Cf. the requirement, mentioned in §1.2, that a semantics be interpretive.

reality.¹⁰

In response to this suspicion, we should distinguish between 1) a representation's possessing a tenseless truth-condition, in the sense of a condition whose obtaining does not vary over time; and 2) a representation's truth-condition being aptly statable without employing tensed vocabulary. 2) is, plausibly, sufficient for 1), but the reverse entailment does not hold. It may be that the only way of stating the truth-condition of a tensed token representation, in a manner which adequately captures' speakers understanding, is by giving another tensed statement. But, it does not follow from this that the truth-condition thereby given will be temporally variable. Rather, it may be that the appropriate verbal expression of the relevant truth-condition will vary depending on the temporal relation of its expression to the time of the token representation. For instance, if on Monday I say, 'Today it is sunny', then the appropriate statement of my utterance's truth-conditions on Monday will be 'What J.B. said is true iff it is sunny today', and on Tuesday will be 'What J.B. said is true iff it was sunny yesterday.'¹¹

This way of giving truth-conditions is motivated by the idea that what I state in saying 'Today it is sunny' is essentially a *dynamic* thought: a thought the grasp of which essentially depends upon a capacity to keep track of events in time. This idea is expressed by Gareth Evans:

¹⁰Compare the device, standard in formal semantics, of including reference to a possible world in the specification of truth-conditions. There is no commitment thereby to a Lewisian picture of possible worlds as forming a vast spatiotemporal array in which the actual world is just the local region of the array that we speakers of the language happen to inhabit. In general, inclusion of semantic indices—such as world-indices—is typically motivated by demands of compositionality, for instance in order to provide an element of the semantic value of a sentence that can get shifted by intensional contexts. It is hard to read any metaphysical implications off such internal, structural considerations.

¹¹Rumfitt 1993 proposes this as the correct form that a Davidsonian theory of truth for indexical languages should take. If, for instance, someone produces an utterance *u* of the sentence-type 'Rumfitt is a fool', then an adequate theory of truth for *u as used by Rumfitt* should issue in the theorem: '*u* is true iff I am a fool'. Rumfitt explains: 'So long...as we understand the theory to be *propounded* in a particular context, there is no indeterminacy in *using* an indexical expression such as "I" to give the Truth-conditions of an utterance...the context of [the theory] determines (among other things) who is speaking [i.e. propounding the theory], and thereby determines what has to be the case for a saying in which "I" is used to be true. This elucidation of the Truth conditions of such reports does deliver the results that we expect. Tony, we are supposing, does say "Rumfitt is a fool", using the name "Rumfitt" to designate me.' Similarly, any interpretive theory of truth (or, more generally, any semantic theory) will be used at a particular time, and so its user can exploit such features of the context in using temporally indexical expressions to express truth-conditions.

...being in the same epistemic state may require different things of us at different times; the changing circumstances force us to change in order to keep hold of a constant reference and a constant thought—we must run to keep still. From this point of view, the acceptance on d_2 of ‘Yesterday was fine’, given an acceptance on d_1 of ‘Today is fine’ can manifest the *persistence* of a belief in just the same way in which acceptance of difference utterances of the same sentence ‘The sun sets in the West’ can. (1985b, p. 309)

Countenancing the possibility of dynamic thoughts means recognising that tense may be used, not to express a condition which is variable over time, but rather to keep track, as time goes by, of something temporally invariant, like a particular event or series of events.

This is an extremely important point. It means that commitment to a tensed metaphysics is not adequately expressed by saying that a complete description of temporal reality must employ tensed expressions. Here, for instance, is Michael Dummett:

Suppose someone who can observe all events which take place in our universe, or some region of it, during some period of time. We may first suppose that he observes them successively, that he cannot choose which events he will next observe but can observe them only in the order in which they take place. Then even if he knows both what he has observed and what he is going to observe, he cannot give a complete description of his observations without the use of temporally token-reflexive expressions. He can give a complete narration of the sequence of events, but there would remain to be answered the question, ‘And which of these events is happening now?...If instead we imagine the observer as able to survey the whole course of events at once...then we can conceive of him as observing a static-dimensional configuration, one dimension of which represents time...It is now clear, however, that what he observes can only be a model of the sequence of events in our three-dimensional space, not that sequence of events itself. (1978a, pp. 354–355)

Dummett’s point is that there is a genuine difference between a dynamic system—a system which is in motion—and a merely static four-dimensional array; the latter can be a model of, but not identified with, the former. But, he claims, this difference is one that can only be expressed through the use of tensed (‘token-reflexive’) expressions. He takes this point to require that the temporal reality described must itself be tensed—motivating McTaggart’s paradoxical conclusion that past, present and future apply equally to all events in time. But once the category of dynamic thoughts is admitted, this step looks like a non-sequitur. It may be that the dynamic nature of time is

something that can be expressed only by means of tensed notions. But if that is because tensed notions express dynamic thoughts, there does not follow any sense in which the reality represented is itself tensed. In particular, it does not follow that the tensed statements used to capture the dynamic nature of reality have truth-conditions whose obtaining varies over time. Thus, the dynamic character of time may be on all fours with the (non-dynamic) nature of space. Perhaps understanding spatiality requires being *in* space, and so deploying egocentric spatial notions. For instance, perhaps the difference between a left- and right-handed golf club is one that can only be grasped at an egocentric level of thought in which ‘left’ and ‘right’ figure as monadic directional predicates.¹² In that respect, a complete description of spatial reality would have to employ egocentric spatial locutions. But maintaining this would not give expression to a bizarre view on which the monadic, egocentric concepts of left and right express intrinsic features of reality.

I have not yet said anything in support of the claim that tensed representations do express dynamic thoughts. The point is rather that recognising the possibility of dynamic thoughts renders more obscure the notion of an intrinsically tensed reality. Two standard ways of formulating a commitment to that picture are, first, that tensed representations have tensed truth-conditions; and, secondly, that no complete description of reality can be given in tenseless terms.¹³ But, as we have seen, the first of these conditions is equivocal between the denial that tensed sentences have cognitively equivalent tenseless counterparts, and the far stronger claim that tensed sentences have truth-conditions whose *obtaining* is variable over time. The first of these can readily be granted; that, however, makes it more obscure what the second comes to, given that we must be able to formulate principles, at some theoretical level, which chain together the truth-conditions of tensed representations produced at different times. As for the idea that there is no complete description of tenseless reality, I have suggested that this may be true if temporal reality can only be represented by means of dynamic thoughts, and that the truth-conditions of dynamic thoughts cannot be given tenselessly. It thus remains unclear what a vindication of the intuition expressed by Difference would require.

¹²The point about left and right goes back to Kant 1992a,b, 2004. It is made in the context of the present issues by Eilan 1997.

¹³For example, see Skow 2015, ch. 2.

2.2 THE A-THEORY AND ITS DISCONTENTS

Someone gripped by the conviction that there is a real difference between past and future might well be unsatisfied with the above discussion in terms of truth-conditions. They might feel that the coordination of the truth of tensed representations over time is a relatively superficial matter; what is really at stake is the purely metaphysical fact that the present represents an absolute boundary: between fixity and openness, reality and unreality, actuality and possibility. The point I have been urging is that this intuition comes under pressure as soon as we notice the systematic way in which the truth-conditions of tensed representations vary over time. Once we recognise that what is past and future varies according to which moment is taken as present, it is no longer clear what it would be for the present to constitute an absolute boundary in the way Difference seems to require.

The key challenge is to find a way of rejecting the move from the systematicity implied by Variation to the picture of an underlying tenseless reality making true tensed representations. The way of accomplishing this is to offer an alternative, positive picture of temporal reality and the passage of time. The required picture is one that might be summarised with the slogan: *Reality changes*. Thus, it might be said, the reason that the same tensed representation-types may have different truth-conditions when tokened at different times is not because those times constitute different perspectives on a tenseless reality. Rather, it is because reality is a different way at those different times; things have changed. A sentence like ‘It is raining in London’ expresses a way for the world to be, *simpliciter*, and the world changes with respect to whether or not it is that way. Just so, ‘My 30th birthday is in the future’, and ‘My 30th birthday is in the past’ represent different ways for the world to be with respect to what is past and what is future; and this, like whether or not it is raining in London, is a respect in which reality changes.

This picture makes tensed representations perspectival in a radical, Nagelian way: it has them expressing facts which are proper to a temporal perspective. Moore expresses this thought as follows:

Colin is speaking at the start of a football match. He says, ‘The match is beginning.’ His representation is true and perspectival. It is from a particular temporal

point of view. The temptation this time is to think that it is true in virtue of a feature of reality that is there only at the start of the match, namely its beginning. Its beginning is not the same as its beginning at 3 p.m. on Saturday, 6 November 1971, even though that is the time (let us say) at which Colin is speaking. The latter is a feature the match never loses...Its *simply* beginning, on the other hand, is there only from the relevant temporal point of view...As before, then, the temptation is to think that Colin's representation is made true by a perspectival fact, the fact that the match is beginning, which obtains only from his then temporal point of view. (Op. cit., p. 44)¹⁴

As before, Moore argues that this notion of a perspectival fact is incoherent. His case here is stronger, however. Whereas in the case of colour it was unclear that we had anything like the plurality of perspectives required to generate a conflict between perspectival facts, this is precisely what we have in the case of different tensed representations tokened at different times. For instance, if on Monday I say 'The nectarine is unripe', and on Tuesday 'The nectarine is ripe' (speaking of the same nectarine), both utterances can be true. But if my utterances have tensed truth-conditions—if, in speaking truly, I am stating a tensed fact—then their joint truth requires both that the nectarine be unripe and that it be ripe. That is a contradiction. In order for the truth of token representations produced at different times not to make conflicting demands on reality, we are obliged to think of tense as a feature of representations, rather than the reality represented. So runs Moore's argument against radically perspectival facts.¹⁵

This line of thought, however, is not conclusive. As Moore acknowledges, it rests on a fundamental premise about the unity of reality: that there is some coherent, total way that things are, of which the various local facts are facets. Elsewhere, Moore urges, in Kantian fashion, that we should see this premise as grounded in a pure intuition of the unity of time:

The unreality of tense, or equivalently the unity of temporal reality, is of a piece with the unity of apperception. We can achieve insight into it through a kind of self-conscious reflection. But we cannot argue for it. Thus if we consider our own

¹⁴Cf. Nagel 1986, 57, n.1.

¹⁵This is, essentially, an argument against tensed representations having tensed truth-conditions. Versions of this argument are to be found in Evans 1985a; Fine 2005; Mellor 1981, 1998. Percival 2002 offers a detailed critical assessment of the argument, and develops a response. Percival's proposed response essentially involves rejecting the conception of time presupposed in saying that tensed representations tokened at different times have different truth-conditions. See n. ¹⁹.

tensed representations and reflect on how we conceive them and what they answer to, then we can come to see them as representations, from different positions within a single world, of that single world—just as, if I consider my own representations (tensed or tenseless) and reflect on how I conceive them and my capacity to produce them, I can come to see them as representations produced by a single self-conscious subject occupying different positions within a single world. (2001, pp. 389–390)

Moore's Kantian intimation of a link between a grasp of the temporal unity of reality and self-consciousness is highly suggestive. These are themes that will be returned to in subsequent chapters. But for now, what this passage brings out nicely is that the unity of reality is precisely what the tense realist may deny. One may doubt whether time constitutes a unified domain, encompassing multiple perspectives, in the way that space does. Indeed, the slogan *Reality changes* encapsulates a view of reality on which it fails to constitute a single, stable totality.

The options for someone who wishes to uphold the reality of tense are helpfully brought out by Kit Fine's reconstruction of McTaggart's argument. Fine urges that the question of the reality of tense must be construed as making explicit reference to a metaphysical concept of reality. The primitive notion is that of a fact's constituting, or belonging to, reality.¹⁶ Equipped with this notion, Fine formulates the McTaggart argument in terms of an inconsistent quartet of principles:

Realism: Reality is constituted (at least, in part) by tensed facts.

Neutrality: No time is privileged, the tensed facts that constitute reality are not oriented towards one time as opposed to another.

Absolutism The constitution of reality is an absolute matter, i.e. not relative to a time or other form of temporal standpoint

Coherence: Reality is not contradictory, it is not constituted by facts with incompatible content. (Fine 2005, p. 271)¹⁷

¹⁶Fine suggests, but does not consistently adopt, a formal regimentation of this notion as a sentential operator \mathcal{R} , to be read as 'In reality, it is the case that _____. I will follow Fine in sticking to the informal locutions.

¹⁷Fine formulates a more sophisticated version of the argument a few pages later using slightly different principles, which does not assume a single, privileged notion of constitution. For easy of exposition I will focus on the simple argument, referring in passing to the sophisticated version where necessary.

Fine gives the argument for their incompatibility as follows:

It follows from Realism that reality is constituted by some tensed fact. There will therefore be some time t at which this fact obtains. Now Neutrality states that reality is not oriented towards one time as opposed to another. So reality will presumably be constituted by similar sorts of tensed facts that obtain at other times (given that there are other times!). We wish to show that it then follows that reality will be constituted by incompatible facts. Now there is no logical guarantee that the facts constituting reality that obtain at t will be incompatible with the facts constituting reality that obtain at other times, since reality might be so boring that the same tensed facts hold at every single time. However, any reasonable view of how temporal reality might be constituted should allow for its being reasonably variegated over time; and presumably it will be then be constituted by incompatible facts, i.e. by facts with incompatible contents. If, for example, it allows for the present fact that I am sitting, then it should also allow for the subsequent fact that I am standing. By Absolutism reality is absolutely constituted by these facts; and this is then contrary to Coherence. (Ibid., p. 272)

The argument implies that one of these principles must be given up. And, indeed, dropping each principles gives rise to a distinctive metaphysics of temporal reality. I will briefly review these in the following paragraphs.

Giving up Realism leads to the orthodox B-theoretic or tenseless view of time—in Fine’s terms, anti-realism about tense. This is the view on which there are no tensed facts, only tensed representations. Tense is indexically, not radically, perspectival.

Views which reject Neutrality Fine dubs ‘standard tense realism’. It is perhaps best to illustrate what standard tense realism involves by focusing on a particularly clear example of it. This can be found in the work of A. N. Prior, which uses the formal resources of tense logic to state a position about the nature of time.

The fundamental idea of Prior’s theory of time, captured by his version of tense logic, is that an ordinary, present-tensed sentence—like ‘The nectarine is mouldy’—states a way for reality to be, *simpliciter*. Past- and future-tensed sentences, by contrast, possess a complexity that shifts their subject-matter away from simply how things are: namely, to how things *were* or *will be*. This complexity is analogous to the way in which modal expressions shift the subject-matter of a sentence from how things are to how they *might have been*, and intensional verbs shift it to how things are *believed [by X] to be*.

The grammatical-logical expression of this idea is that the past and future tenses

function as sentential adverbial modifiers. Prior puts it like this:

...putting a verb into the past or future tense is exactly the same sort of thing as adding an adverb to the sentence. 'I *was* having my breakfast' is related to 'I am having my breakfast' in exactly the same way as 'I am *allegedly* having my breakfast' is related to it, and it is only an historical accident that we generally form the past tense by modifying the present tense, e.g. by changing 'am' to 'was,' rather than taking on an adverb. ('Changes in Events and Changes in Things,' p. 13)

The formal expression of this in tense logic is that past and future are regimented as one-place sentence-level operators, like the \Box and \Diamond of modal logic. By contrast, there is no need for a present-tense operator: unembedded, simple sentence just state simply how things are.

The metaphysical reflection of this point is, as Prior elsewhere puts it, that 'These two concepts [of the real and the preset] are closely connected; indeed on my view they are one and the same concept, and the present simply *is* the real considered in relation to two particular species of unreality, namely the past and the future.' (1972, p. 320) The point here is that reality just *is* present reality. There are still truths about how things were and how things are; but these must be understood, not as truths about other regions in a temporal domain of which the present is a distinguished location, but rather about a particular aspect of present reality, namely its past- or future-tensed aspect.

There is a clear sense in which this view of time is one that gives up Neutrality, albeit one that is not most happily stated by saying that a particular time is privileged as the present. On Prior's picture, any reference to a domain of times has ultimately to be understood as an abstraction from the tensed truths which hold in the present (that is to say, from the tensed truths which hold *simpliciter*.) Prior's way of doing this is to identify times with maximal propositions of tensed truths. To say that a proposition φ holds at a time i , then, is just to say both that i was or will be the case, and that i implies φ .¹⁸ More generally, the idea is that the notion of something's being the case at a time t is captured in terms of a conditional: if t were present, such-and-such would be the case. By contrast, the present truths are just those which are categorically the case. Fine summarises this by saying that, on theories of tense that reject Neutrality, other times exist as *hypothetical* realities. (Ibid., p. 279)

¹⁸I discuss Prior's tense logic in more detail in chapter 6.

According to standard tense realism, at least according to Prior, there is no sense in which tense is perspectival. The perspective of the present just is the ‘perspective’ of reality itself. Accordingly, standard tense realism effectively denies that there is a multiplicity of temporal perspectives.¹⁹

The sense in which what is past and what is future changes is just: what is past used to be future; and what is future will be past. These principles can be expressed by means of tense-logical theorems involving nested applications of tense operators. But their holding is, as it were, internal to a single, present-tensed reality.

Fine argues that standard tense realism has difficulty accounting for the passage of time. The problem is, put picturesquely, that the facts which, according to the standard tense realist, constitute reality, are ‘frozen’ on a single present. He puts the problem like this:

The standard realist faces a general difficulty. For suppose we ask: given a complete tenseless description of reality, then what does he need to add to the description to render it complete by his own lights? The answer is that he need add nothing beyond the fact that a given time *t* is present, since everything else of tense-theoretic interest will follow from this fact and the tenseless facts.²⁰ But then how could this solitary ‘dynamic’ fact, in addition to the static facts that the anti-realist is willing to accept, be sufficient to account for the passage of time? (Ibid., p. 287)

The problem seems to be that the passage of time requires that there be not merely *one* distinguished present, but many—passage involves the transfer of presentness, whatever that may be, from one time to another. Any metaphysics that drops Neutrality, then, can at best offer a kind of snapshot or cross-section of passage. A fully general and comprehensive account of a changing reality, by contrast, will have to somehow encompass a multiplicity of present-tensed realities. And this means retaining Neutrality.

This is not offered as a decisive argument. I think we should allow that standard realism is at least a distinctive and coherent position, albeit one that may not offer a

¹⁹Similarly, the problem of providing truth-conditions for token representations produced at different times does not arise for the standard realist. This is the key move in Percival 2002’s response to Evans and Mellor. See n. 15.

²⁰On Prior’s standard realism, which takes times to be abstractions from tensed truths, the ‘tenseless facts’ will be just those that are always true. In particular, there will be series of conditionals of the form *Always, if i, then...*, where *i* is an instant-proposition.

sufficiently robust account of passage to satisfy all parties. I shall have more to say about tense logic, and the assimilation of tense to modality, in chapter 6, which will ultimately count against standard realism. For now I will continue the review of Fine's varieties of tense realism.

Fine terms views that retain Neutrality and Realism 'non-standard tense realism'. The result of dropping Absolutism is what he calls 'external relativism'; and rejecting Coherence leads to 'fragmentalism'.

On external relativist views, the constitution of reality is a matter which is *irreducibly* relative to a time. The external relativist employs a primitive notion of reality being constituted-at- t by a fact f , where, crucially, this does not give way to reality's being (absolutely) constituted by the time-relativised fact f -at- t .²¹ For instance, reality is constituted-at-Jan 1, 2019 by the fact *It is raining in London*, where this is not explained in terms of the absolute obtaining of the fact *It is raining in London on Jan 1, 2019*. External relativism thus contrast to the orthodox anti-realist view on which something's being past, present or future is 'internally' relative to a time of reference.

The second non-standard tense realism, fragmentalism, is one on which there just are flat-out contradictions between the facts that constitute reality. For instance, the facts *It is raining in London* and *It is not raining in London* may both constitute reality. Reality is not just a pile of contradictory facts, though. The facts somehow arrange themselves into 'fragments', with each fragment corresponding, intuitively, to the more standard notion of how things are at a time.

Fine's taxonomy of tense realist metaphysics cuts across the more traditional varieties of the A-theory of time. Presentism, the most orthodox version of the A-theory, is most naturally taken to be a form of standard tense realism.²² By contrast, other versions of the A-theory, most notably the 'growing block' and 'moving spotlight' theories, may be taken to admit of standard or non-standard interpretations.²³

²¹In the official regimentation, the operator \mathcal{R} is replaced by one containing an argument-place for times \mathcal{R}_t .

²²Prior is, on any reasonable interpretation, a presentist. More recent defenders of presentism are Bigelow 1996; Bourne 2006; Markosian 2004.

²³In what follows I focus just on growing block theories. Moving spotlight theories that are readily understood as standard tense realisms include Sullivan 2012b and Deasy 2015. Meanwhile Lipman 2018 and Skow 2015 present moving spotlight theories that are explicitly inspired by Fine's fragmentalism.

In the case of the growing block view, a standard reading would be one on which, in the tensed facts that constitute present reality, there is asymmetry with respect to the range of the existential quantifier occurring within past- and future-tensed contexts. That is: it is never the case that there will be something that is not now identical with something; whereas it is sometimes the case that there is or was something that was not always identical with something.²⁴ This view—which Braddon-Mitchell (2013) calls the ‘pure tensed growing block’ or PTGB—can be understood as effectively a combination of Prior-style standard realism with a thesis about quantification.²⁵ To that extent, it is subject to the same concerns Fine expresses about whether it adequately captures the passage of time.

Once we consider non-standard versions of the growing block view, things become more complex. Perhaps the most natural way of understanding the growing block view is as an attempt to combine the four-dimensional manifold described by physics with a tense-theoretic account of how the manifold grows.²⁶ The challenge which confronts anyone who wants to combine these pictures is how to make sense of the growth of the block. The familiar objection is that the view requires a mysterious second time-dimension in which the block can grow.²⁷ Fine’s non-standard realism offers more than one way for the growing block theorist to meet this challenge. An external relativist might take the facts which compose reality relative to a time to include (tensed) facts concerning the ‘size’ of the block—for instance, facts like *t is the last moment of time*, and *It will be that t is not the last moment in time*. On the other hand, a fragmentalist might take the facts about each differently-sized block to cohere into moment-like fragments of temporal reality. The chief difference is that the external relativist invokes some kind of super-time series in order provide indices for the relativisation of facts about the size of the block, whereas the fragmentalist does not.

The principal difficulty with such views is that it is hard not to lose one’s grip on the reality, or non-perspectival nature, of the distinction between past and future. Take the

²⁴Formulations of this kind are to be found in Correia and Rosenkranz 2013, 2018.

²⁵Deng 2017 suggests that there is also an external relativist interpretation of PTGB available. I will ignore this complication for now.

²⁶This is the kind of picture suggested by Earman 2008. However it is not clear it applies to Tooley 1997’s growing block view in quite the same way.

²⁷For example, Sider 2001, ch. 2 presses this objection forcefully.

external relativist version. In order to employ the relevant notion of relative constitution, there must be some series of times to which different sets of facts about the ‘size’ of the block are relativised. We can ask about the relation of this time-series to the ‘internal’ time of each block. Suppose it is distinct. Then it looks difficult to identify the block with temporal reality: the block has given way to an ultimate, eternally existing, super-time series. It is very hard not to see the constitution of the block-facts relative to moments in this series as reducing to facts about the size of the block relative to positions in the super-series. The external relativist position then becomes difficult to distinguish from an anti-realist position with a time-relative notion of existence on which, if something exists at t , then it exists at all $t' > t$.

On the other hand, suppose they are identified. Then growth of the block amounts to the fact that, relative to each time t , t is the last moment of time. But this just looks equivalent to saying that, always, the present time is the last moment in time (there is never a time later than the present.) And now this position looks very hard to distinguish from the standard realist interpretation of the growing block theory. This can be seen from the fact that someone endorsing such a theory is committed to affirming that now is the last moment in time. There is no super-time series to which they can appeal to explain the future-directed growth of the block. Of course they will also be able to say that, in the future, there will be moments in time which do not now exist; hence, now will not always be the last moment in time. But this is exactly the kind of thing that a standard growing block theorist will say; affirming tensed facts about the future size of the block does not express a commitment to Neutrality.

Now consider a fragmentalist growing block. Since there is no need to relativise truths about the block to a time, there is no need for any kind of superordinate time-series. However, there is now a question to be asked about what ensures that the different, inconsistent facts hang together in the right way. Fine says very little about this. He states that ‘One would...expect there to be various substantive “rules of coherence” concerning the conditions under which a set of facts would be coherent and the way in which the coherence of one set of facts might constrain the coherence of another set.’ (p. 281) But the content of these rules, and the basis of their reality-organising status, remains wholly mysterious. In articulating any such rules, it is almost irresistible to

want to avail oneself of the idea that the fragments constitute a series, with the relations between the facts that make up each fragment constrained by order-theoretic relations among the fragments. But understanding the organisation of the fragments this way just invites reinterpretation in terms of external realism: instead of a network of fragmentary, incoherent but absolutely reality-constituting facts, one imagines the facts obtaining ‘at’ a fragment or ‘in’ a fragment. And this just seems like a notational variant on constitution relative to a time. So fragmentalism collapses into external relativism, which in turn collapses into either anti-realism or standard realism.

The general problem with both forms of non-standard realism is that it is exceedingly difficult to give any positive content to the key primitive notions—constitution relative to a time, or arrangement into fragments—without reverting to the ideology of some more orthodox theory. In order to retain Neutrality, the non-standard realist has to have some notion of a comprehensive, all-encompassing reality to which the various temporal perspectives belong. But almost any conceivable way of explaining the relation of the various temporally perspectival realities to the more comprehensive reality threatens to lead to the ‘internal’ relativism of anti-realism. Alternatively, one may attempt to do away with the idea of an all-encompassing reality containing the multiple tensed realities. But then it becomes impossible to maintain that there really is a multiplicity of fragments, and we end up back with standard realism.

Fine is sensitive to this general difficulty, stating it thus:

One might say that über-reality ‘manifests itself’ in the form of the particular realities, that it becomes ‘alive’ or ‘vivid’ through the particular realities obtaining. Each particular reality presents itself as the whole of reality. It creates the illusion, if you like, that there are no further facts, even though there are many such realities and each is equally real. But it should be acknowledged that these remarks merely gesture in the direction of a certain idea and that, if we have here a viable conception of a pluralistic universe, then none of the usual models for making sense of it will apply. (p. 283)

This is, essentially, just the original tension between the ideas that distinctions of tense are real, but nevertheless vary. Non-standard realism seeks to capture both of these features by combining an ‘internal’ level in which past, present and future exclude one another, with an ‘external’ level on which the various tensed realities coexist. The diffi-

culty Fine acknowledges is how to keep both levels in view without one collapsing into the other—leading either to standard realism or to anti-realism. The only way to hold them apart is to keep the faith with the heavy-duty metaphysical notions of fact and reality-constitution required to articulate each position. But, equally, this may lead one to doubt that a distinctive position has really been staked out.

It is especially pertinent to consider this general difficulty for non-standard realism in connection with the possibility raised earlier, that temporal reality can only be described by means of dynamic thoughts. The key point about that possibility was that, even if the canonical way of expressing the temporal facts changes over time, that does not imply the facts themselves do. Now the non-standard realist positions have reality being constituted by different facts at different times, or by inconsistent facts quarantined into different fragments. There will be certain systematic links across fragments between facts: for instance, if that fact *The nectarine is ripe* belongs to one fragment, then we should expect there to be the fact *The nectarine will be ripe* to belong to another, ‘earlier’ fragment.²⁸ Grouping these facts together across fragments, what is the basis for regarding them as distinct? In general, what is the basis for regarding the fragments as distinct sets of facts, rather than different perspectives on the same underlying facts; similarly, what is the basis regarding the facts that obtain relative to one time as distinct from those that obtain relative to another, rather than as different perspectives on the same underlying set of non-relatively obtaining facts? An initially compelling answer to this question is that if the facts are to be identified, they are tenseless facts, and so must be describable in tenseless terms—and there is no tenseless description of the facts available that adequately captures the dynamic nature of time. But if the facts are essentially expressible only by means of dynamic thoughts, then we should not expect to be able to express them tenselessly. Nevertheless, we have no reason not to identify them: splitting up reality into different chunks, each corresponding to a different way of stating the facts, looks like unchecked scholasticism.

To paraphrase a famous remark of Nagel’s, it is not so much that we have decisive reason to reject non-standard tense realism, but that we do not understand what it

²⁸Of course to say that the fragment is earlier cannot be taken to imply that the fragments do form a time-series; rather, the idea is just that the facts of each fragment are so coordinated that any present tensed fact in one has answering to it a future tensed fact in the other, and so on.

would be for it to be true. (1979c, p. 176) I have argued that, given the possibility of dynamic thoughts, we have to understand the A-theory in terms of a properly metaphysical notion of fact, and of the constitution of reality. But the problem is how to get at the relevant notion of a fact other than via the truth-conditions of sentences that express them.

I think we should conclude that our grip on the notion of a temporal fact, and our assessment of the various forms of non-standard tense realism on offer, will ultimately be disciplined by intuitions about what the passage of time amounts to. If one has a sufficiently robust idea of passage, one may take that to support a picture of a changing reality, one which can then be articulated in terms of a manifold of tensed facts, perhaps along the lines of one of Fine's non-standard tense realisms. But we should not expect the relevant notion of a fact, or of reality-constitution, to be explicable independently of the phenomenon of passage they are invoked to explain.

Someone with a robust metaphysics of passage, then, may rely on such a metaphysics in order to explain what the commitment expressed by Difference amounts to, and whether it can be made consistent with Variation. Ultimately, I wish to remain neutral about the prospects for such a project. In the remainder of this chapter, though, I want to raise the possibility of a more deflationary understanding of Difference, one which employs the embedding strategy to make sense of the perspective from which the difference between past and future shows up. This will prepare the ground for the idea, to be developed in subsequent chapters, of agency as involving a temporal perspective on events.

2.3 PERSPECTIVE ON AND IN TIME

The alternative strategy I want to put forward in the remainder of this chapter is to focus once more on the sense in which tense, and in particular the difference between past and future, is perspectival. Rather than assuming we have a firm prior grip on the notion of a temporal fact, we might instead proceed by asking what kind of structure in the world we need to appeal to in order to make sense of the perspective from which we employ tensed notions.

We can understand the tension between Difference and Variation, rather than as making a conflicting pair of demands on the metaphysics of time, as a tension that exists within the perspective we occupy on the temporal world. We can then approach metaphysical questions by asking what structure a consistent and satisfactory account of that perspective commits us to in the world.

We might articulate the tension as follows: We think of the past very differently from how we think of the future. We think of the past as something to be grasped, understood, or come to terms with; whereas we think about the future as something to be made or brought about. To put it slightly differently, we are passive with respect to the past and active with respect to the future. At the same time, our understanding of time exhibits a certain systematicity. This comes out especially clearly in our thinking about the past. We have a firm sense of the past as real, in a way that is not exhausted by its traces in the present. An aspect of this sense of reality is plausibly the appreciation that past objects and events were once available to be encountered in just the way that present entities now are. One's understanding of something existing in the past is in this way dependent on one's understanding of that very same thing having once existed in the present.²⁹ Hence the difference represented between past and future cannot be just a difference in the *kind of information* represented, the past being distinguished as containing one type of event and the future as containing another. Someone who thought of the past as comprising its own, proprietary kind of entity would not yet have mastered the concept of the past.

The dependent character of thought about the past thus seems to require the idea of time as affording a multiplicity of perspectives. The problem is that temporal experience does not seem to provide us with anything like the required notion of perspective. Even if temporal experience has an envelope-like character, encompassing a horizon of awareness with a duration of a few seconds, there is nothing like the perspectival structure of the experience of space on offer. Rather, we just experience things as happening.

²⁹This may seem like a platitude. But note that it is denied by, for instance, Russell's (2001 (orig. 1912)) theory of memory, on which 'pastly' is just a primitive way in which an object of awareness may be directly given in experience. This means denying what Campbell 1994, ch. 7 calls the 'stepwise' character of memory, the dependence of memory knowledge on an earlier, more direct mode of awareness. As Campbell argues, this position becomes much more problematic once tries to combine it with the idea that awareness of mind-independent objects is subject to spatiotemporal constraints.

Unlike with space, temporal experience does not seem to provide us with access to a unified domain encompassing multiple points of view. Indeed, experience—specifically, agential experience—seems to present us with a radically *disunified* domain, in which the future is fundamentally different from the past.³⁰

The temptation here is to think that we have two very different, perhaps inconsistent, ways of thinking about time.³¹ There is an immersed and unreflective level of thought, perhaps connected especially with agency. At this level of thought, past and future are given a primitive and irreducible meaning, articulated in terms of the susceptibility of future events to intervention and the inaccessibility of past ones. At the same time, we are capable of reasoning about time in a more detached, theoretical way. At this level of thought we understand time as a linear order of events, and use the tenses to express indexical information about the time of an event in relation to a time of utterance or evaluation.

The problem with this approach, I think, is that it fails to do justice to the way in which the ‘engaged’ and ‘detached’ stances are mutually supportive. The way we understand the difference between past and future in agency and deliberation involves a fairly systematic appreciation of the structure of time and causality. Conversely, our objective conception of the temporal world is supported, fundamentally, by a conception of

³⁰One crucial way in which temporal experience does not seem to be perspectival in the way that spatial, and specifically visual, experience does is underlined in a discussion by Matthew Soteriou. He remarks the visual experience seems to involve an awareness of a region of space, where ‘one can say something, at least very roughly, about the extent and shape of the region of space (delimited by one’s sensory limitations) that one is visually aware of.’ By contrast, ‘when we are consciously aware of an interval of time, of limited extent, within which we seem to perceive objects and events, it doesn’t seem to us as though we are aware of that interval of time from a point in time that is distinct from it—e.g. from a point in time that falls within that interval of time...Introspectively, it doesn’t seem to one as though one can mark out one’s own temporal location as distinct from the temporal location of whatever it is that one is perceptually aware of. It doesn’t seem to one as though one is perceptually aware of an entity occupying some temporal location that is distinct from the temporal location *from* which one is perceptually aware of it.’ (2013, p. 131). A consequence of these features of the phenomenology of temporal experience is that experience does not seem to provide us with the materials for a conception of the domain of times as affording a multiplicity of perspectives. This is because temporal experience does not provide for the idea of a single phenomenon being experience *from* multiple perspectives, in the way that spatial experience does.

³¹This kind of split is explicitly endorsed in McCormack and Hoerl Forthcoming’s recent proposal that adult temporal thought involves both an ‘updating system’ and a ‘reasoning system’, and that apparently paradoxical features of our understanding of time can be seen to arise from the interaction between the two systems. I discuss Hoerl and McCormack’s view in chapter 5.

ourselves and our agency as belonging to it.

This mutual dependence of reflective and immersive, or engaged and detached, ways of thinking about time perhaps comes out most clearly in our concept of the past. We think of the past as distinguished from the future, but at the same time I have stressed that the dependent character of our thinking about the past means we cannot be thinking of it as a peculiar and autonomous domain. Rather, we must be thinking of the past as just a part of a global temporal framework. The challenge is to see how our thinking about the past could have the kind of systematicity required to support this conception, without undermining the difference between our thinking about the past and the future.

I think we can see this difficulty as arising from the assumption that manifest time has the form of a theory: some relatively well-defined body of general beliefs about the structure of time, whose reality-reflecting status we can enquire into. We do exhibit different patterns of thought and concern with respect to past and future events. But it is far from clear that this is because we believe past events to have one property and future events to have another. It may be that, insofar as we are inclined to accept as platitudes certain apparent general truths about time—‘You can’t change the past’; or ‘The future isn’t settled’—these rather reflect attempts to make sense of the messiness and complexity of the temporal world as we experience it.

In his engaging book *The Anthropology of Time* (1992), Alfred Gell argues that anthropologists have tended to attribute to the societies they study overly exotic theories of the structure of time: that time is cyclical, or that ancestral time is a distinct and independent order from that marked by the agricultural calendar, and so on. His argument is that such interpretations read too much explicit theorising into the ways in which society and culture is organised in time, and the various timekeeping and calendrical systems which support those forms of organisation. It is, I think, important to ward against similar mistakes in reflecting on ‘our’ manifest image of time. Craig Callender, for instance, characterises manifest time as involving a number of general theses about the nature of time, including that ‘now’ picks out a unique simultaneity slice of the entire the universe:

Snap your fingers. That snap picks out a present event for you, but you also assume this present extends indefinitely. If there are aliens living on some extra-

solar planet, then their present is the same as yours when you snapped your fingers.
(2017, p. 10)

Contrary to what Callender says, it is far from obvious that anything like this assumption is presupposed by ordinary tensed thinking. We normally use the tenses only in relation to events in our fairly immediate surroundings. The application of such notions to remote regions of space may simply be indeterminate. Entertaining the idea of events in a far-off galaxy is a highly theoretical, reflective achievement; in thinking such thoughts, it may be simply indeterminate how to go on in applying one's everyday patterns of tensed thinking. Indeed we might speculate that the belief in an indefinitely extendible, universal now is itself a sophisticated conceptual innovation—albeit a mistaken one—facilitated by the development of rapid communication and co-ordinated timekeeping systems.

In general, the articulation of manifest time in terms of certain highly general beliefs about the nature of time looks like a suspect philosophical attempt to saddle ordinary thinking with greater systematicity and articulacy than it in fact possesses. It is notable that as soon as one tries reflectively to articulate one's naïve beliefs about time in any remotely systematic way, one very quickly becomes confused, and inclined to say things that seem inconsistent—that only one moment is present, but in another way every moment is present; that events are constantly moving forward into the future, but in another way constantly receding into the past; that the past is real, but in another way unreal; and so on. These observations should make us sceptical that there is any such thing as 'our' naïve theory of time.

Even if that is so, it is not thereby impossible to say anything systematic about manifest time and the type of temporal perspective it embodies. What we need to recognise is that there may be a complex interplay between the structure of a subject's experience, and the way they conceptualise their experience and make sense of its relation to a more encompassing reality. This is where the embedding strategy introduced in the previous chapter can be especially valuable. The idea of the embedding conception was, first, that a subject's perspective is to be articulated partly in terms of the structure of world around them, and the relations they stand in to those bits of worldly structure, rather than just in terms of the content of the subject's beliefs; and, secondly, that fitting the

subject's experiences into an encompassing conception of reality is principally a matter of seeing how those experiences arise from objective features of the world and the subject's relations to those features. What we need to understand is how our manifest image of the time is the upshot of embedding our specific actions and experiences in a more comprehensive picture of the temporal world; and how, as changeable beings embedded in the temporal order, our general conception of time is anchored to the particular episodes that make up our route through the world.

The version of the embedding strategy I will pursue in the following chapters accords a central place to the idea that our experience and thought relates us to temporal particulars—to concrete unrepeatable events. An idea I shall unpack in more detail in chapters 5 and 6 is that the kind of systematic understanding of time involved in possession of the concept of the past is supported by this element of particularity in our thinking. At the same time, thinking about events, and in particular about the events of our own lives, is grounded in the actual temporal and causal structure of those events themselves. We can only to a limited extent abstract our thought from these structures in which which are inextricably entangled. Consequently, our embedding of our actions and experiences in a more comprehensive conception of the world is limited in certain ways by the structure of those processes themselves. Specifically, what this means is that there is a certain way of thinking about ourselves as elements of the causal order—a way of thinking about our own agency that is reflective, but not wholly detached or disengaged—that we deploy only in relation to the past. This introduces a fundamental asymmetry into our thinking about time, but not one based in explicit beliefs about the difference between the past and the future. It is an asymmetry in our thinking itself, rather than in what we think.

This characterisation of the project is so far extremely abstract. In the following chapter I will turn to a different way of articulating the difference in our thinking about the past and the future, one grounded in the structure of deliberation. Seeing what is missing in this approach will then set the stage for the development of my proposed version of the embedding strategy.

3 | ASYMMETRIES OF CAUSALITY AND DELIB- ERATION

The previous chapter discussed, in rather abstract terms, the question of the reality of the difference between the past and future. I argued that a tensed metaphysics, on which the distinction of past and future corresponds to a difference in reality, has trouble in accounting for the variation and systematicity in our thinking about time. Trying to accommodate this systematicity, by reintroducing a variety of temporal perspectives corresponding to different divisions of past, present and future, threatens to undermine the distinctiveness of the tense realist position. It is thus not clear what we might be after in wanting to affirm the reality of the difference between past and future.

At the end of the last chapter I suggested that, rather than understanding the apparent difference between the past and the future in terms of our having certain general beliefs about the structure of time, we might instead see it as primarily a function of our nature as agents, and the temporal orientation of our agency. The most salient context in which our thinking about the past and the future differ is deliberation. As Aristotle notes in book VI of the *Nicomachean Ethics*, ‘nothing that is past is an object of choice, e.g. no one chooses to have sacked Troy; for no one deliberates about the past, but about what is future and capable of being otherwise, while what is past is not capable of not having taken place; hence Agathon is right in saying, “Of this one thing is even god deprived/ To make what has been done not to have happened”’. (2000, 1139b) We might therefore seek to explain our different treatment of the past and future in a more deflationary manner, in terms of the intrinsic orientation of the process of deliberation.

This raises the question why we deliberate over the future and not the past, if not

due to a prior belief in the fixity of the past. One answer invokes our understanding of causality. We deliberate only over those things we can causally effect; in addition, we have a firm and widespread belief that causes always precede their effects. Hence there is only a point in deliberating over events which are subsequent to the time of deliberation. Since the future is just the time subsequent to the present, this is equivalent to saying that there is a point to deliberating over the future but not the past.

But this in turn raises questions, first, of the basis of our belief in the direction of causality; and, secondly, why causal structure should constrain deliberation. It is a basic part of our causal thinking that causes are ‘effective strategies’ (Cartwright 1979): that is, whatever the causal relation is, it should be such that, if C causes E, then bringing about C (or not-C) is a way of bringing about (not-)E.¹ The question is why the notion of cause should have this connotation, and why it should always go in the direction earlier-to-later.

One response to these questions takes the content our causal notions to be fixed, ultimately, by the role they play in guiding deliberation. On this approach, there is a level of description of agency and deliberation at which causal notions drop out. Typically this level comprises an account of practical rationality along Bayesian decision-theoretic lines, and the (non-causal) correlations which provide evidential inputs to an agent—perhaps an ideal agent—whose deliberation is characterised by the decision theory. The content and truth of causal beliefs is ultimately explained with reference to this level of description. There is a broad analogy here with explanations of colour discourse in terms of properties of perceivers’ visual systems, without reference to the colours of things. Menzies and Price, for instance, state that ‘the concept of causation is to be explained by relation to our experience as agents in the same way that the concept of colour is to be explained by relation to our experience as observers.’ (1993, p. 193) This position might be taken to express the idea that causal beliefs are perspectival, in a sense that implies that causal structure is not a feature of reality. It is thus attractive to those moved by the fact that fundamental physics describes the world in time-symmetric, non-causal terms.²

¹This formulation is similar to that of Price 1991; Price and Menzies 1993. A slightly different statement of the mutual constraint between causal structure and rational deliberation is given in Fernandes 2017.

²This line of thought is expressed in Russell 1912’s polemic against causal notions. It appears explicitly

I shall ultimately be arguing that this conception of cause is inadequate, as there is more to the use of our notion of cause than its role in guiding deliberation. In particular, in this chapter and the next I shall be urging a need to recognise a kind understanding we have of ourselves and our actions which is indeed causal, but essentially retrospective. This retrospective standpoint makes use of causal notions that do not simply track the information that was available to the agent at the time of deciding. There is thus more to the notion of cause than we can discern just by attending to its role in rational deliberation; insofar as understanding of causal structure does constrain deliberation, it does so by appeal to an imagined, subsequent retrospective standpoint. Recognising this retrospective dimension to causal thought will then be argued to inform a correct account of the temporal perspective of agency.

In the first place, though, I want to motivate a sceptical line of thought concerning the notion of cause and its relation to the fixity of the past. The problem is to identify some objective feature of time and causality which implies the fixity of the past, but does not also imply fatalism. I will begin in section 5.1 by considering Michael Dummett's discussion of the possibility of bringing about the past. Dummett argues that any argument for the fixity of the past is precisely analogous to a seemingly fallacious fatalistic argument about the future. Disarming the fatalistic argument thus seems necessarily to loosen our grip on the idea that the past is indeed fixed. In section 5.2 I further pursue this line of thought by arguing that there is no intrinsic absurdity in the possibility of backward causation and time travel, at least from the perspective of a deliberating agent.

Section 5.3 recasts Dummett's dilemma in the setting of Newcomb's problem, and the debate between Causal and Evidential Decision Theory. I argue that, if we take seriously the sceptical challenge posed by Dummett's dilemma, it is hard to justify the causalist's reasoning entirely on the basis of the instrumental function of deliberation. In section 5.4 I recast some of these ideas in the context of the interventionist approach to causality developed by Woodward (2003) and others. I argue that the cases in which causalist deliberation is problematic are just those cases in which one's actions do not have the status of interventions. This will be so, in particular, when causal thinking is applied to one's own agency. In the final section I conclude that we need to acknow-

as a motivation for agency-based theories of causation in Fernandes 2017; Price 1997.

ledge a self-conscious use of causal thinking that does not merely track information available to deliberating agents in planning their interventions, but involves taking a non-deliberative, reflective stance towards oneself and one's actions. I argue moreover that this stance is specifically associated with a retrospective perspective on one's actions.

3.1 DUMMETT ON BRINGING ABOUT THE PAST

Why do we take the past to be closed to deliberation? One might take the fixity of the past to consist just in the fact that there are determinate truths about what has happened.

One way of articulating this thought is as follows. Take some event *E* such that I am concerned over whether or not it has occurred. One might reason, either *E* has or occurred or it has not. If *E* has occurred, then it will have occurred whether or not I take some proposed course of action. If it has not occurred, then it will not have occurred whether or not I take the proposed course of action. Either way, my actions make no difference to the occurrence or non-occurrence of the event in question.

Michael Dummett points out that this argument is precisely analogous to a fatalistic argument about the future. Suppose that the event in question lies in the future. We can reason analogously, either the event will occur or it won't. If it will, then it will if I do or don't take some proposed action; if it won't, then again it will whether or not I take the action.

Quasi-formally, here is the general schema of both forms of the argument:

- (1) Assume: *E*
- (2) If *A*, then *E*; and if not-*A*, then *E* (From 1)
- (3) So, *A* is ineffective in preventing *E* (From 2)
- (4) Assume: Not-*E*
- (5) If *A*, then not-*E*; and if not-*A*, then not-*E* (From 4)
- (6) So, *A* is superfluous in preventing *E* (From 5)
- (7) *E* or not-*E* (LEM)
- (8) So, *A* is either ineffective or superfluous in preventing *E* (From 1,3,4,6,7)

(9) So A makes no difference to whether E occurs

(From 8)

The future-directed version of argument is obviously fallacious; but it is not easy to pinpoint exactly how. And it is more difficult still to identify a flaw in the fatalistic argument that does not equally apply to the argument about the past. Strikingly, neither argument appears to rely on specifically temporal considerations. They each appeal just to the Law of Excluded Middle, and the principle that, under the supposition of p , the conditional *if q , then p* is correct.³ The latter step of course fails for subjunctive or counterfactual conditionals. But the conditionals involved in the argument are indicative conditionals, for which the principle is compelling: we are supposing that the actual world is one in which p holds. Under that supposition, p still holds in the actual world whatever other q we suppose about it—even if the combination of p and q is highly improbable.⁴

Dummett's diagnosis of the fallacy is that the move (under the supposition that E will occur) from the conditional 'if I do not do A, then E will occur' to the fatalistic conclusion that my actually doing A will be superfluous in bringing about E, is invalid. More precisely, it is invalid under any interpretation of the conditional which also validates the inference $p \Rightarrow \text{if } q, \text{ then } p$. As he observes, for the inference from 'if I do not do A, then E will occur' to the fatalistic conclusion to be compelling, the conditional ought to be incompatible with the one that negates its consequent: 'if I do not do A, then E will not occur'. If both conditionals can be true, then the first cannot reasonably be regarded as expressing the independence of E from my action A. But, if the principle $p \Rightarrow \text{if } q, \text{ then } p$ is valid, then these conditionals are not incompatible: they may both be true as long as the antecedent is false. That is, they may both be true precisely in the case in which I will in fact do A, thereby bringing about E.⁵

³The argument can also be construed as an application of an unrestricted version of the decision-theoretic principle of Dominance. Cf. Ahmed 2018a.

⁴Moreover, it is worth noting that on what are probably the two main rival theories of indicative conditionals—the material conditional theory, and the suppositional theory—the principle $p \Rightarrow \text{if } q, \text{ then } p$ is valid. See Edgington 2014 for an overview.

⁵This rebuttal of the argument only works on the assumption that the conditional is a material conditional. On a suppositional theory of conditionals such as Edgington's, *if q , then p* follows from p , but is not consistent with *if q , then not- p* . That is, if $P(p) \approx 1$, then $P(p|q) \approx 1$ as long as $P(q) > 0$ and otherwise undefined, whereas $P(p|q) + P(\sim p|q) = 1$ as long as $P(q) > 0$ and otherwise undefined. So this analysis

But now, Dummett points out, the same fault can be identified in the argument about the past. From the fact that, if E has occurred, then it will have occurred whether or not I do A, we cannot legitimately conclude that, if I actually do A, my doing of A is of no relevance to the prior occurrence of E. For it could be precisely my doing of A that explains the occurrence of E. Just as, in the future case, the conditionals in question can hold and it can still be the case that, given that E will actually happen, it will happen as a result of my doing A, it can equally be that E has occurred, but only because I am to bring it about by doing A.

Whether or not one accepts Dummett's specific diagnosis of the fatalistic fallacy, we are thus presented with a dilemma. Either we should accept his conclusion that there is no intrinsic absurdity in bringing about the past, or else accept the analogous future argument and embrace its fatalistic conclusion. Both horns would, to a considerable extent, undermine our ordinary ways of thinking about the difference between the past and future. In order to overcome the dilemma, we need to find some difference between the arguments, or some way of repairing one that does not apply to the other.

Dummett's discussion does not actually rule out that there might be a conditional which validates the argument for the past and not for the future. If, for example, the conditional is a Lewisian 'non-backtracking' counterfactual, (cf. Lewis 1987b) then the inference from p to *if q , then p* will be correct just as long as p concerns something temporally prior to q . Of course it is a further question whether the inference is strictly *valid*, for some appropriately stringent standard of validity; but even if not, it is still the case that, in the actual world, whenever p is true, $q \Box \rightarrow p$ is also true as long as p concerns something earlier than q . And if the pair of counterfactuals 'if I were not to do A, E will have occurred', and 'if I were to do A, E will have occurred' is true, there is

of conditionals calls for a different resolution of the fallacy.

I think that what we should say is that the last step of the fatalist's argument by cases is invalid. That is, we can allow that, *on the supposition that E will occur*, if you do A it will be unnecessary for bringing about E, since, conditional on E occurring, the probability of E is the same as the probability of E given A, namely 1. Trivially, $P(p|p) = P(p|p, q) = 1$. Similarly, by the same reasoning, on the supposition that E will not occur, if you do A it will be ineffective in bringing about E, since, conditional on E not occurring, the probability of not-E is the same as the probability of not-E given A, namely 1. But we cannot infer from these two cases that whether A occurs is irrelevant to whether E occurs. This is just point, familiar from Simpson's Paradox, that, for any two dependent events, it is always possible to construct an partition of states of the world such that conditioning on any member of the partition makes the two events independent.

a far more compelling case for the conclusion that my doing A is irrelevant to the prior occurrence of E.

But the appeal to non-backtracking counterfactuals, or any other time-asymmetric conditional, is controversial at this point. Dummett is raising a question about the basis of our differential treatment of the past and future in deliberation. It seems highly likely that any tendency we have to hold the past, but not the future, fixed when evaluating certain conditionals is a reflection of some more basic asymmetry in our thinking, rather than an adequate explanation of it. Lewis holds that the asymmetry of counterfactual dependence is based on a metric of comparative similarity between possible worlds, itself not intrinsically temporally asymmetric. But Lewis's criteria of similarity are reverse-engineered specifically to falsify backtracking counterfactuals; and it is in any case not clear that they can be applied clearly and precisely to yield the correct range of counterfactual judgments.⁶

Similarly, one might think that our different treatment of the past, and our use of time-asymmetric conditionals, is based on an appreciation of asymmetric causal structure. But this too may be argued to beg the question against Dummett. Part of the point of the dilemma is to probe the basis of our belief in asymmetric causal structure. It should not be assumed, at least at this stage, that we have a firm grip on the asymmetry of causation independently of our deployment of conditional information in practical reasoning.

Dummett himself concludes that our differential treatment of the past and future is based simply on the fact that we never encounter certain correlations which would justify a policy of acting to bring about the past. However, he argues that we can easily enough imagine circumstances which would justify such a policy. Suppose that there is a strong correlation between my doing A, and E happening at some time prior. Clearly this alone does not establish anything like a backward causal connection of the kind that would be required to think of my doing A as efficacious in bringing about E. But now suppose then whenever E does not occur (at the time at which one would expect it to occur), if I attempt to do A, one of two things almost always happens: either something unexpected prevents me; or else I successfully do A, but it subsequently turns out that I

⁶For critical discussion of these points see Horwich 1989, ch. 10; Woodward 2003, pp. 137–145.

was mistaken in thinking that E had not occurred.

Here is a variant of an example Dummett uses. Imagine an agrarian society who subsist on a particular crop as their staple food. After being harvested, the crop has to be shut up in a dark storehouse for six months and left to ferment. The fermentation is a slow, delicate process that happens gradually over the period, and success is not guaranteed. It cannot be checked on as opening the storehouse before the end of the six months will ruin the crop. However, there is a formal ceremony that is normally performed by the official in charge of the crop on the day the storehouse is opened, and there is a strong correlation between correct performance of the ceremony, and that year's crop being properly fermented. The question Dummett asks is, could the members of this society be justified in regarding the performance of the ceremony as a means of bringing about the successful fermentation of the crop over the previous six months?

Dummett asks what would happen if we introduce into the situation some means of inspecting the condition of the crop in advance of the official opening. Perhaps we can put a device into the storehouse which detects, without altering, the atmospheric conditions, and is wired up to a meter outside, so it can be used reliably to find out about the condition of the crop. Now we ask the official to perform a series of experiments: Whenever the meter says that the crop has failed, they perform the ceremony anyway.⁷ If in all or most of these cases, they succeed in performing the ceremony, then they would be forced to conclude that the ceremony is not a way of bringing about a successful crop after all. On the other hand, suppose that, on most of these occasions, either of the following happens: i) something, unexpectedly, goes wrong with the performance of the ceremony; or ii) it emerges there was some malfunction with the meter and the crop is fine. Dummett urges that, in this case, we could not criticise their continuing practice of performing the ceremony to ensure a successful crop.

If the above results of the experiment were observed, it would undermine the reliability of the meter as a way of finding out whether the fermentation process had occurred.

⁷This is the 'bilking experiment', intended as an argument against backward causation, discussed by Black 1955; Flew 1954; Mackie 1966; for some more recent critical discussion of the argument, cf. Garrett 2014; Roache 2009. Horwich (op. cit., chs. 6 & 7) persuasively argues that what bilking argument shows is not that local backward causation is impossible, but that it would almost always involve inexplicable coincidences. This is also acknowledged by Dummett in his later (1993) discussion.

The only good way of finding out is to open the storehouse at the end of the period, with or without performing the ceremony. Dummett concludes that the crucial feature of our thinking about the past is that we have ways of finding out what happened that we think are evidentially insensitive to our intentions. Memory and records are paradigm examples of such ways of learning about the past. Knowledge of the future, by contrast, is in general not independent of one's intentions. If, on the other hand, we had genuine memory-like foreknowledge of the future, we might often regard future events to be fixed in the way we do the past.⁸

The question how to interpret Dummett's case can be factorised into two subsidiary questions: 1) Would members of the society be justified in regarding the correct performance of the ceremony as a cause of successful fermentation of the crop?—more generally, is an exploitable correlation sufficient for the existence of a causal connection?; and 2) Is belief in a causal connection required for it to be rational for the official to perform the ceremony with the intention of ensuring a successful crop?—that is, must deliberation exploit only causal connections? These are both instances of more general questions: 1) can we coherently describe circumstances in which one would be entitled in believing in a backward causal connection?; and 2) is it always irrational to act for the sake of some result, in the absence of a causal connection from one's action to the result?

It is worth distinguishing these two questions, as answers to them can be disciplined by different motivations, which may come into tension. One may, for instance, answer 1) negatively on the grounds of some other constraints on the notion of cause. Perhaps it is a matter of definition that causes precede their effects; or, less crudely, perhaps it is a necessary condition on causation that there be some spatiotemporally continuous mechanism connecting cause and effect, and there are principled reasons why no such mechanism can be located in Dummett's case.⁹ But if one denies causation on these grounds, it becomes harder to see why the absence of a causal link should matter for the rationality of the official's performance of the ceremony. Why should the presence or absence of mechanisms affect whether something is an effective way of bringing about

⁸Garrett 2016 argues that this condition of intention-independence is not quite right, and offers some refinements.

⁹See Dowe 2000; Salmon 1984 for mechanism-based theories of causation.

one's goals, other than the fact that a causal mechanism is a sign of a correlation that can be exploited in means-end reasoning? On the other hand, one might follow Price and others (op. cit) in taking the content of causal claims to be fixed by their role in deliberation, or at least ideally rational deliberation. In that case, one will likely answer 2) in the affirmative: the direction of causation and deliberation will not come apart. But then there is no longer any basis for denying that, in a case like Dummett's, the connection is causal. The next two sections discuss each of these questions in more detail.

3.2 BACKWARD CAUSATION, CAUSAL LOOPS AND TIME TRAVEL

One might legitimately ask whether Dummett's case is really possible as described, and whether it would indeed provide evidence of backward causation—rather than one in which an unusual set of correlations give rise to an illusory sense of control over the past.

One question is whether, in general, backwards causation is conceptually possible, or whether it is consistent with universal background conditions. On the crude Humean view that causes by definition precede their effects, having backward causation is impossible in the way that an unmarried bachelor is. But this is not an especially interesting way of ruling out backward causation, since we can always ask whether something exactly like the causal relation other than its temporal direction could hold between a later event and an earlier one.

A somewhat more sophisticated view is that temporal order is defined as following the direction of causation (E.g. Mellor 1998; Reichenbach 1956.) But this means that temporal order is that of the *dominant* direction of causation; it does not necessarily rule out that there may be local exceptions. It may be that the supposition of backward causation gives rise to absurdities (of the kind to be discussed below), but this will not be a matter of definition.¹⁰

On this latter view, or on any view where temporal order is conceptually independent from the direction of causation, we can sensibly ask why the dominant direction of

¹⁰This is in fact Mellor's position, cf. *ibid.*, chs. 10 & 11.

causation is from earlier to later, and whether exceptions are possible. This question can arise both for reductive views on which the causal relation is identified with certain regularities, and nonreductive ones on which the causal relation is left unanalysed. Any plausible nonreductive view of causation will still allow that the causal facts constrain which regularities can occur, even if they are not wholly determined by those regularities. It is thus legitimate to ask about the relationship between causal structure and background non-causal regularities, and whether those background conditions are consistent with local instances of backward causation.

It is a striking fact that direction of causation is not a matter of physical law. When we look at the fundamental laws of physics, we not only find no mention of cause, but differential equations which distinguish earlier and later no more than they distinguish left and right—they are ‘time-reversal invariant.’¹¹ So at the very least, backward causation is not ruled out by physics at the fundamental level.

On the other hand, the observable universe is rife with *de facto* asymmetries: processes whose time-reversal is consistent with physical law, but never actually occurs. Chief amongst these is the asymmetry of entropy, stated by the Second Law of Thermodynamics: a splash of milk quickly disperses throughout a cup of tea, but tea and milk never spontaneously separate. In addition is the asymmetry of radiation: in Popper’s (1956) famous example, ripples spread outwards from a disturbance on the surface of a pond, but never converge on a point. Higher-level irreversible processes, like biological processes, are related in a complicated and obscure way to these more basic *de facto* asymmetries.

Asymmetries of entropy and radiation are normally thought to reduce to stochastic mechanical principles, such that the underlying asymmetries themselves are merely statistical, and neither nomic nor causal. And this may in turn be thought to explain the direction of causation.¹² However statistical principles alone are not enough, since they only explain why high-entropy states are more probable than low-entropy ones; this point by itself does not explain why low-entropy states always precede high-entropy

¹¹This point is stressed by Russell 1912. See Albert 2000, ch. 1 for a particularly clear statement of time-reversal invariance.

¹²The most thoroughgoing philosophical attempt to reduce the direction of time and causation to statistical relations is Reichenbach 1956.

ones, or indeed) why they occur at all. Statistical explanations of *de facto* asymmetries must thus appeal to some further background conditions: for example, facts about the initial conditions of the universe.¹³ Whatever these further conditions are, it is largely unsettled whether they are consistent with the possibility of local exceptions to the direction of *de facto* irreversible macro processes. Insofar as the direction of causality is constrained just by that of irreversible processes, then, it is just unclear whether backward causality is a live physical possibility.

One notable potential instance of backward causation is the possibility of backward time travel. Time travel in science fiction stories generally involves the use of time travelling devices, so that there is a clear point when the backward journey begins and ends. But a (marginally) more realistic scenario involves smooth closed timelike curves, where an object or person makes a journey beginning and ending at the same spacetime point. Reichenbach vividly describes the following possibility:

Some day you meet a man who claims that you are his earlier self. He can give you complete information about your present condition and might even tell you precisely what you are thinking. He also predicts your distant future, in which you will some day be in his position and meet your earlier self. Of course you would think the man insane and walk on. Your companion on [your world-line] agrees with you. The stranger goes his way with a knowing smile; you lose sight of him as well as of your companion...and forget about both of them. Years later you meet a younger man whom you suddenly recognise as your earlier self. You tell him verbatim what the older man had told you; he doesn't believe you and thinks you are insane. This time you are the one that leaves with a knowing smile. You also see your former companion again, exactly as old as he was when you last saw him. However, he denies any acquaintance with you and agrees with your younger self that you must be insane. After this encounter, however, you walk along with him. Your younger self disappears from sight from then on you lead a normal life. (1958, pp. 141–142)

Unlike stories involving time machines, closed timelike curves do not require abrupt spatial and material discontinuities in the world-lines of time-travelling objects.¹⁴ Moreover, closed timelike curves are not only possible to describe cogently; in 1955, Kurt

¹³Philosophical discussion of these issues include Albert 2000; Horwich 1985; Price 1986.

¹⁴Dummett 1993 discusses some of these discontinuities, such as the fact that a device sent back in time would abruptly appear at the time and place of its destination, requiring the air molecules in the region to have suddenly rushed out to create a vacuum nanoseconds prior to its arrival.

Gödel proved that there are possible solutions to the field equations of General Relativity that imply their existence. (Having said that, Reichenbach's example is unrealistic in that the amount of energy required for a single body to traverse a closed timelike curve in Gödelian spacetime renders any such scenario technologically impossible.) Backward causation is involved here in the sense that some event is a cause of an effect which occurs earlier on a timelike curve—for instance, the old man's speech and the young man's memory of it. But in another sense there is no violation of the dominant direction of causation, since, unlike in Dummett's case, there is also a timelike curve running in the earlier-later direction from cause to effect.

Even if unusual causal structures are consistent with global conditions, though, there remains a question whether we can coherently describe a scenario involving backward causation. One might have doubts about whether descriptions of backward causation such as Dummett's really hold up to scrutiny.

It is certainly true that Dummett's case is underspecified, and, arguably, spelling it out in more detail gives rise to near-absurdities. In particular, there will have to be a series of inexplicable coincidences when we consider the finer structure of the purported backward causal connection. The fermentation of the crop will presumably have, or appear to have, an ordinary history of prior causes when it does occur, which causes are not present when it fails to occur (for instance, the atmospheric conditions in the storehouse.) How are we to suppose the performance of the opening ceremony relates to these apparent prior causes? Does it affect the fermentation process only through its atmospheric causes, or are they independent? If the first, then how does it bring about *those* causes (which presumably each have their own prior histories)? If the second, then why do the appropriate atmospheric conditions always coincide with the performance of the ceremony?

Moreover, in Dummett's case it is stipulated that whenever the fermentation has gone wrong (whenever it has actually gone wrong—not merely whenever the meter indicates that it has) and the official tries to perform the ceremony, they fail for some apparently unrelated reason. What explains this coincidence? Either it is again inexplicable, or there must be a hidden causal connection running from the failure of the process to whatever disrupts the ceremony, undermining the supposed evidence for the

hypothesis that the ceremony is a cause of the fermentation.

Related problems arise for time travel and closed causal loops. A familiar paradox of time-travel is that it seems both possible and impossible that a time traveller could kill one of their ancestors. It is impossible because doing so would make it the case that the time traveller never existed, contradicting the supposition that they travelled back in time to kill the ancestor. But on the other hand, killing the ancestor seems to be something it is in their power to do, an instance of ordinary agential freedom.

The standard response to the supposed paradox is that the sense in which the traveller is free to kill their ancestor is not one which implies that there is a genuine *global* possibility, consistent with the full range of conditions which enabled them to get into that position, in which they actually do so.¹⁵ There are presumably many things which we are free to do but which, when all the causal factors preceding our choice are taken into account, it is impossible that we should actually do. (Arguably, determinism implies that this is the case for all our possible actions other than those we actually take.) If there is any conflict here then it is a purely general one between freedom and determination, not a special one that arises for backward time travel.

But even if there is no real paradox, there are some oddities. Any attempts by time travellers to kill their ancestors will fail; we can be certain of this because the supposition that they succeed is inconsistent. But this is no explanation of why the attempts must fail. Rather, attempts always fail for some incidental, unrelated reason. The explanations of individual failures are, crucially, not evidence of any incapacity on the part of the time traveller. They are just ordinary instances of someone failing, for whatever reason, to do something that was nevertheless within their power. But in that case, the repeated failure of time travellers to kill their ancestors would be an inexplicable coincidence. This is essentially the same point as the one made above, that the consistent failure in Dummett's scenario of attempts to perform the ceremony when the fermentation is known not to have occurred would be inexplicable.

The fact that time travel and backward causation would involve inexplicable coin-

¹⁵This is the response of Lewis 1987c. Lewis connects this response with a specific thesis about the meaning of 'can': to say that something can happen is to say that its happening is compossible with a certain, contextually specified, set of facts. This semantic thesis is not required for the resolution of the apparent paradox.

cidences means that many of our normal ways of thinking about time would be disrupted. We are used to explaining apparent coincidences by reference to common causes. Specifically, when two events are correlated but neither is a cause of the other, we can typically find an earlier event whose occurrence renders the later events probabilistically independent.¹⁶ Such patterns of explanation are, plausibly, a central connotation of ordinary causal thinking, and in particular are implicated in the deployment of singular causal claims to explain particular chains of events. Singular causal relations are normally thought to involve an asymmetry of explanation: causes explain their effects, but not vice versa; and coincidences can be explained by identifying or positing an earlier event as a common cause. On a grand scale, this practice of explanation from earlier to later implicitly supposes that all apparent coincidences can be ultimately referred back to universal initial conditions. (And, as Lewis (op. cit., p. 74) points out, the occurrence of *those* conditions is a contingency whose inexplicability we have to live with.) We can note this intimate connection between causal thinking and explanation without thereby attempting to reduce causal claims to some model or rather of explanation (for example, subsumption under general laws.)

The strangeness of nonstandard causal structures thus lies at least partly in the fact that, in such cases, familiar habits of explanation are not usable, leaving us with brute unexplained facts in places we are not accustomed to find them. But it may be that our explanatory habits are adapted to fit our world, where we never in fact encounter the correlational patterns that would require us to violate the standards of those practices.¹⁷ Time travel and backward causation, whilst they may be metaphysically and even nomically possible, thus tend to involve features that would disrupt our normal ways of making sense of the world. As such, it does seem that backward causation would be seriously jarring, if not downright inconsistent, with some fairly deeply embedded features of ordinary thinking.

However, these oddities all lie on the side of explanation. Dummett's discussion, by contrast, is supposed to test the idea that there is any intrinsic asymmetry in *deliberation*,

¹⁶This idea is at the heart of Reichenbach's (1956) account of causal asymmetry.

¹⁷This is roughly Horwich's (op. cit., ch. 7) attitude to Gödelian timelike curves. He argues that there is no real conflict between the possibility of Gödel spacetime and our usual explanatory principles since, as mentioned above, traversing a closed timelike curve is technologically impossible, even in principle. Thus, even in Gödel spacetime, the problematically inexplicable coincidences could never in fact occur.

by presenting a situation in which it would apparently not be absurd or irrational for someone to act for the sake of bringing it about that something happened. Even if it turns out that spelling out the details of the situation leads to explanatory absurdities, those absurdities may emerge only on reflection, rather than from the point of view of the deliberating agent. In general, it is not clear how exactly the explanatory function of causal thinking is related to its role in constraining deliberation.

One kind case in which backward causation does have the potential to introduce absurdities into deliberation is when an agent can be certain about what they will or will not do. For instance, a moderately reflective time traveller can be certain that any ancestricidal attempts will fail; and this might be argued to imply that it would therefore be absurd, or even impossible. The thought here would be that deliberation and intentional action are subject to the rational constraint that one cannot decide to do what one knows one will not do. But if this thought is correct, it at most means that there is a genuine sense in which a (rational, reflective) time traveller cannot kill their ancestor, or at any rate cannot do so intentionally.¹⁸ The fact that certain courses of attempted action are not rationally open to time travellers does not render their deliberative situation intrinsically absurd.

More puzzling are cases where an agent, for whatever reason, can be certain that they will do something voluntarily, or make a certain decision, rather than the merely negative fact that they will not successfully perform an action of a certain kind. Williams (1992, p. 138) discusses the prospect of what he calls ‘immediate fatalism’: the attitude, while deliberating, that the outcome of that very deliberation is both inevitable and certain. Immediate fatalism is absurd, he argues, because it undermines its own supposition that the outcome of the deliberation will be a *decision*. Recognising that the outcome of one’s deliberation is inevitable would be a passive acceptance of the outcome, bringing deliberation to a close.¹⁹ But then one could not be deciding after

¹⁸This is argued by Fernandes forthcoming, who goes on to connect our ordinary lack of certainty regarding our future actions with our sense of the future as open. Cf. also Rennick 2015.

¹⁹Note that this is just to say deliberation is inconsistent with certainty about what one will choose *when that certainty has some independent basis outside one’s deliberation*. Being certain, or at least highly confident, that one will choose a certain option when practical reason speaks decisively in its favour is not immediate fatalism. Some decisions are just no-brainers. Thus one can consistently hold immediate fatalism to be psychologically impossible while admitting the cogency of Joyce 1999’s arguments against Levi 1993’s doctrine that deliberation requires suspending all credence in one’s options.

all—contradicting the idea that one knew in advance that one would decide.

I think we should agree with Williams that immediate fatalism is a psychological impossibility. But it is an extreme case, and it is unclear that it ever need arise in the context of time travel and backward causation (unlike the supernatural setting of Greek tragedy which is the focus of Williams's discussion.) Whilst backward causation may sometimes require that an agent be certain of some outcome, it is harder to see how an agent could know with certainty that the outcome depended specifically on a decision or voluntary action of theirs. Moreover, even if the agent did know this, they may not know exactly when and how they would take the action or decision in question. Immediate fatalism only arises if an agent is certain, while deliberating, what the outcome of that deliberation will be. And it is hard to see how this particular absurdity is required by the possibility of backward causation.

In summary, whilst it may be that we never encounter, and perhaps even could never encounter, situations involving backward causal connections, it does not seem that there is anything intrinsically absurd about the possibility from the point of view of a deliberating agent. This suggests that the appropriate response to Dummett's dilemma is to be located in contingent facts about the structure of the world, and the kinds of exploitable causal connections we actually encounter, rather than the internal structure of deliberation. But this in turn is open to the challenge that deliberation need not necessarily follow the direction of causation.

3.3 NEWCOMB: CAUSATION AND EVIDENCE

Discussion of the relation between causation and rational decision has focused on what is essentially a variant on Dummett's case, namely Newcomb's problem.²⁰ The problem goes like this: there are two boxes, one transparent and one opaque. The transparent one visibly contains £100; you are told that the opaque one contains either £1,000,000 or nothing. You have the option either of taking just the opaque box, or both boxes. However, prior to choosing, a highly reliable Predictor has carried out extensive psychometric tests on you to predict which option you will go for. If she predicts you will

²⁰The original presentation of the problem is due to Nozick 1969.

take both boxes, the opaque box will be empty; if she predicts you will take just the opaque box, it will contain the £1,000,000.

Unlike Dummett's example, Newcomb's problem is not typically discussed in connection with questions about the metaphysical or conceptual possibility of backward causation; the usual context is rather the normative question what decision procedure is optimally rational in cases where evidential and causal dependencies diverge, on the assumption that in Newcomb they do diverge. Hence, the problem is not typically framed as one in which a backward causal connection is a live epistemic possibility. But we may wish to describe the setup more sparsely, as one in which the Predictor's predictions are strongly correlated with your actual choice and you have no access to information about her prediction other than by choosing,²¹ leaving unspecified the method by which she makes her prediction. It is then open to a theorist to redescribe the case as one where one's choice causes the prior prediction.

This strategy will be favoured by someone who wishes to explain causal structure in terms of patterns of ideal deliberation, understood ultimately in evidential terms. The advantage of explaining causal structure in this way is that it explains why causal structure should matter to deliberation in the way we normally take it to. Alison Fernandes states this explanatory goal clearly:

It is not enough to simply claim that the relations picked out as causal are relevant to our practical lives, or merely stipulate that the relations are ones of control or counterfactual dependence. One has to actually show that the relations picked out as causal or counterfactual are those that matter to our practical lives. There are many relations we might have picked out as causal, and many ways we might have evaluated the relevant counterfactuals. Why are these the right ones? How do they matter to us? Can we show by some independent means that we should pick out these relations as causal? (Fernandes 2017, p. 687)

One way of answering this challenge is to take causal notions fundamentally to track facts about what an ideal deliberator, characterised in purely evidential, non-causal terms, would do. Thus Price describes cause as an 'expert function': 'causal dependence should be regarded as an analyst-expert about the conditional credences required

²¹This second condition distinguishes a genuine Newcomb problem from so-called 'medical Newcomb problems'—on which more below—where it one can have prior introspective access to the beliefs and desires that are correlated with the states of the world relevant to the outcome of one's choice.

by an evidential decision maker.' (2012, p. 509) The view is intended by analogy with the subjectivist's understanding of chance as the subjective probability an ideally rational thinker assigns to an event. On such a view, the following set of stipulations about the Newcomb case are jointly inconsistent: 1) The problem involves a genuine decision; one is free to decide either way, and one's decision is a cause of what one does; 2) One's decision remains a source of evidence about the contents of the opaque box even when all other available evidence is taken into account: there is no information that even an ideally rational agent could acquire, prior to choosing, that would render the contents of the box evidentially independent of their choice; 3) One's choice is not a cause of the contents of the box. Holding this view thus commits one to rejecting the supposition that Newcomb problems do not involve backward causation, when other features are held fixed.

On the other hand, someone who takes it that causal and evidential dependencies, even for ideally rational thinkers, can come apart, is owing an explanation of why in such cases we should go with the causal rather than the evidential dependencies. Arif Ahmed puts the challenge like this:

...why should its causal efficacy be what settles the value of an option? No binary relation is more poorly understood than whichever causal one [the causal decision theorist's conditional] is meant to denote. It is intolerably mysterious to say without further grounds that we should...act on the basis of its extension. (Op. cit., p. 113)

To bring this point into focus, consider the idea discussed in the previous section that there is an intimate connection between cause and explanation. Perhaps this demand that a cause be explanatory is articulated in terms of a demand for a mediating mechanism, or a covering law. One might then resist the causal redescription of the Newcomb scenario on the grounds that one's choice does not explain the distribution of prizes: rather, both should ultimately be explained in terms of a postulated prior mental state, albeit a state that one has no introspective access to. But then there is a question how this explanatory structure bears on the question what it is rational to do. It is not obvious why it should. I want the world to be such that there is money in the opaque box, and am confident it will be if I refuse the extra £100. Why should I care whether the presence of the money is explained by my choice, or whether both are explained by some prior

event? It begins to look as though there are two, fundamentally different perspectives here: the perspective of deliberation, and the perspective of causal explanation. Why should we expect the notion of cause as deployed in both perspectives to align?

Leaving aside this background dialectic, let us revert to the standard description of the Newcomb case as involving no backward causal connection, bearing in mind that the consistency of this description may in principle be denied. Understood in the standard way, Newcomb's problem is the focus of disagreement between two rival theories of rational decision, Evidential and Causal Decision Theory. Evidential Decision Theory (as it has subsequently come to be known), as developed by Savage (1954) and Jeffrey (1983), has at its heart the idea that a rational agent maximises the expected value of their actions, where an action's expected value is the sum of the values an agent attaches to its various possible *outcomes* (pairings of actions with states of the world) weighted according to their subjective probabilities. Jeffrey's decision theory computes the value of an action using Bayesian conditional probabilities, specifically the probability of each state of the world conditional on the act in question. Here is a Jeffrey-style formula for evidential expected value (EEV):

$$(EEV): EEV(A) = \sum_{1 \leq i \leq n} Pr(S_i|A) V(A \& S_i)$$

In the Newcomb case, EDT gives the recommendation that you should take just the opaque box. Here is the payoff matrix:

	Predict one (S_1)	Predict two (S_2)
Choose one (A_1)	£1,000,000	£0
Choose two (A_2)	£1,000,100	£100

Suppose the Predictor is 90 per cent accurate. The conditional probabilities are $Pr(S_1|A_1) = Pr(S_2|A_2) = 0.9$ and $Pr(S_1|A_2) = Pr(S_2|A_1) = 0.1$. The evidential expected value of taking one box is thus $EV(A_1) = Pr(S_1|A_1) V(S_1 \& A_1) + Pr(S_2|A_1) V(S_2 \& A_1) = 0.9 \times £1,000,000 + 0.1 \times £0 = £900,000$, and the evidential expected value of taking both boxes is $EV(A_2) = Pr(S_1|A_2) V(S_1 \& A_2) + Pr(S_2|A_2) V(S_2 \& A_2) = 0.1 \times £1,000,100 + 0.9 \times £100 = £100,100$. So you should take one box.

Causal Decision Theory (CDT) aims to capture the idea that a rational agent chooses the act which is likely to cause the best result, rather than being evidence of the best res-

ult. One way of putting it is that CDT is concerned with the efficacy of actions, rather than their auspiciousness. The *causal expected utility* of an action is a probabilistically weighted sum of the values of the possible outcomes of an action given its likely causal consequences. This means excluding from the utility calculation any evidential dependencies that do not reflect causal relations.

There are various different approaches to implementing this. One way, due to David Lewis (1987a), is to revert to unconditional probabilities, including as requirement on the proper formulation of a problem that the probabilities be defined over a partition of *causal hypotheses* $\{K_1 \dots K_n\}$: full specifications of how ‘all the things [the agent] cares about do and do not depend on his present actions’ (p. 313). ‘Dependence’ here is causal dependence, which for Lewis is equivalent to counterfactual dependence. Thus his causal hypotheses are effectively large and complex conjunctions of counterfactuals.²² A Lewis-style formula for causal expected utility (CEU) is thus simply:

$$(\text{CEU}): CEU(A) = \sum_{1 \leq i \leq n} Pr(K_i) V(A \& K_i)^{23}$$

Given the assumption that the distribution of prizes is causally independent of your choice, the two relevant dependency hypotheses may be taken just to be the same as EDT’s S_1 and S_2 , namely that there is money in both boxes or there is just money in the transparent box. It can then easily be seen that, since in either case the value of taking both boxes exceeds that of taking one by £100, one should take both boxes no matter one’s prior probability distribution over the dependence hypotheses.²⁴

²²Since it is stipulated that *all* relevant causal information is included in the dependency hypotheses, we may assume that each hypothesis is itself causally (though, of course, not necessarily probabilistically) independent of the agent’s actions, since if it were not, it would then fail to include some relevant causal information about how the causal facts depend on the agent’s actions, contradicting the supposition that it is maximally specific. Eells 1982, ch. 5 provides useful discussion.

²³The use of unconditional probabilities means that Lewis’s version of CDT is not *partition-independent*: it requires the partition of K_i s to meet certain constraints, and there is no internal guarantee that formulations of the same choice situation with different partitions will deliver the same recommendations. This means that the theory is far less flexible in its application to a wide range of choice situations. See Joyce 2000 for development of criticism along these lines. The challenge is to secure partition-invariance without relying on Bayesian conditional probabilities and so introducing spurious non-causal dependencies. This is one of the central tasks of Joyce 1999.

²⁴This is an instance of the principle of Dominance. The problem was originally presented by Nozick as involving a conflict between Utility Maximisation and Dominance; in subsequent discussion it is more usually understood as involving a conflict between maximising EEV and CEU.

By far the orthodox answer in the large literature stemming from Newcomb's problem is that taking both boxes is the rational choice. (Indeed, in many discussions it is principally presented as a technical challenge for how to rule out 'spurious' non-causal dependencies from a formal decision theory.) At the time of choice, the Predictor has already determined what the contents of the opaque box are; since her actions lie in the past, you no longer have any way of affecting them. Whether the opaque box is empty or full, you will get £100 more if you take both—so you should take both. Of course, your reasoning this way is good evidence that the opaque box will be empty. This is unfortunate, but, at the time of choosing, there is nothing you can do about it; you make the best of the bad situation by taking both boxes. Taking one box for the sake of its auspiciousness would be to adopt, in Lewis's memorable phrase, 'an irrational policy of managing the news'.

This judgment can be made vivid by comparing it to more realistic cases, so-called 'medical' Newcomb problems. Suppose it is known that, rather than smoking being a cause of lung disease, there is a gene which causes both lung disease and a tendency to smoke. Assuming that you would find smoking moderately pleasant and that you strongly want to avoid lung disease, should you smoke or not?²⁵ EDT seems to say you should not, since smoking makes it statistically more likely that you have the gene and hence that you will develop lung disease. But intuitively this is the wrong answer: smoking makes no *causal* difference to whether you have the gene. Either you already have it or you don't. So you should go ahead and smoke if you want to. This of course mirrors the reasoning rehearsed above in favour of taking both boxes in the Newcomb problem—your choice makes no difference to the prizes, it is just a sign of how they have already been distributed. So you should take both boxes.

The evidentialist has some room for manoeuvre here. Defenders of EDT typically argue that it is in fact able to give the same recommendations as CDT in Newcomb problems. The basic idea is that the fact of one's smoking is mediated by prior mental states—for instance, an inclination to smoke—that are just as strongly correlated with the gene as is actually smoking. The prior states 'screen off' the evidential dependency of the gene on smoking. If one has some kind of introspective access to such information,

²⁵This and other similar examples are offered in Gibbard and Harper 1978.

it is no longer the case that smoking raises the probability of one's having the gene. So, correctly described, medical Newcomb problems do not involve a divergence between the causal and evidential facts.²⁶

However, it is definitive of a genuine Newcomb problem that this strategy is not available. It is built into the case that one has no idea how the Predictor makes her prediction, and so there is no possibility of the evidential import of one's choice being screened off by prior introspectible states. Thus, as long as one holds fixed the assumption that there is no backward causation, any genuine Newcomb problem is necessarily a case where the evidential and causal dependencies diverge, and hence so do EDT and CDT. (Of course, if we redescribe the problem as involving a backward causal connection, then EDT and CDT converge once again, this time on the recommendation of one-boxing.)

The evidentialist will argue that, due to this difference, in a genuine Newcomb case you should follow the evidence and take the one box. The most prominent objection to CDT's typical recommendation is the so-called 'Why ain'tcha rich?' argument: one-boxers will end up much better off than two-boxers, and this is entirely predictable in advance. But CDT recommends being a two-boxer. How can this be rational?²⁷ Moreover, unlike the medical Newcomb problems, where it seemed clear that the presence of the gene was not up to you, whether to be a one- or two-boxer is precisely the kind of thing that is in one's power. Insofar as the causalist recommends being a two-boxer, they recommend doing something they know will make you worse off.

The standard causalist response is that, at the time of choosing, the distribution of prizes is not up to you. It is true that you would be better off if you were the sort of agent who takes one box. If you could take a pill, or be hypnotised, or whatever, in order to become a one-boxer, this is what CDT would recommend—but only, of course, if you take the pill before the prediction (or are playing an iterated version of the game.) But this is consistent with the fact that, once the prizes are distributed and you have to choose, it is always better to take both boxes and receive the additional £100. If you

²⁶This is the so-called 'tickle defence'. Different implementations of it are offered in Eells 1982; Horwich 1985; Jeffrey 1981; Price 1986. Ahmed 2010 argues that this strategy is fundamentally unstable, arguing we need to embrace a notion of what is chosen by the agent that may include at least some of the past history, as well as the consequences of their decision.

²⁷For discussion of the objection, see Ahmed 2018b; Ahmed and Price 2012; Joyce 2007, 2018; Lewis 1981.

took the pill to make you a one-boxer, you would do better than if you had stayed a two-boxer. But even then, when it came to choosing, you would still do worse than if you had taken both boxes after having taken the pill. (Perhaps taking the pill ensures that this will not happen—but that does not make it irrational, only unlikely.) Thus when it comes to the point of choosing, the one-boxer and the two-boxer face different situations: the opaque box contains a million pounds for the two-boxer, and is empty for the one-boxer. The two-boxer can then argue that they made the best of their (bad) situation, whereas the one-boxer did not make the best of theirs.

Joyce states this rejoinder forcefully:

It is part of the definition of a Newcomb problem that the decision maker must believe that what she does will *not* affect what the psychologist had predicted.²⁸ Letting M stand for “The \$1,000,000 is in the bank [i.e., the opaque box]”, A be taking the money [in the transparent box], and B be refusing it, this requires Irene [the one-boxer] to assign a high subjective probability to both of the following:

$M \supset [(A \Box \rightarrow M) \& (B \Box \rightarrow M)]$	If the \$1,000,000 is in the bank, then it would still be there were I to do A and it would still be there were I to do B .
$\neg M \supset [(A \Box \rightarrow \neg M) \& (B \Box \rightarrow \neg M)]$	If the \$1,000,000 is in the bank, then it would not be there were I to do A . and it would not be there were I to do B . ²⁹

If Irene really does believe these things, then she can give no good answer to Rachel’s question [why she didn’t take the money in the transparent box] since, having gotten the \$1,000,000, she must believe that she would have gotten it whatever she did, and thus that she would have done better had she taken the \$1,000 [i.e. the contents of the transparent box]. So, while she may feel superior to Rachel for having won the million, Irene must admit that her choice was not a wise one when compared to *her own* alternatives. The “If you’re so smart why ain’t you rich” defense does nothing to let Irene off the hook; she made an irrational choice that cost her \$1,000. (Joyce 1999, p. 153)

The core of the causalist response is thus that, at the time of choice, the one-boxer and the two-boxer are in different causal situations. Since you cannot choose which causal

²⁸Setting aside, as before, the complaint that this description of the situation is not consistent.

²⁹Note that these conditionals are just counterfactual versions of those used in Dummett’s argument against bringing about the past.

situation you are in, it is a mistake to infer from the causalist's predictably worse result that they chose less rationally.

However, the evidentialist can reply that this simply assumes the causalist's conception of choice. In choosing one box, you are certainly not *causing* it to be the case that, at the time when the Predictor made her prediction, you were such as to (predictably) be going to take the one box. But this need not imply that it is not among your options as something you can choose. Of course, if you were not in that condition, then you will not actually choose it—but only a fatalist thinks that 'can' implies 'will'. What makes it a matter for deliberation, according to the evidentialist, is just that there is no possibility of independent reflection on the question whether you were at the earlier time in a condition of being such as to take the one box; it is in the nature of the case that one has access to those facts only through deliberation. The evidentialist thus denies that the one-boxer and two-boxer are in different situations, when the situation is conceived from the point of view of the agent.

All of our choices—apart possibly from some rare acts of pure libertarian freedom—have some causal preconditions which are rendered more probable by our so choosing. We may have little reason to care about these in ordinary instrumental reasoning, but in Newcomb they are precisely the states that matter to our success. The question is what grounds the causalist has for refusing to regard these as objects of choice just as much as the effects of one's decision are.

The causalist's refusal to recognise these as options is connected with their use of causal counterfactuals, like the ones in the passage from Joyce above. The dependency judgments of concern to the causalist are counterfactuals which hold fixed the causal history of the act when evaluating its likely consequences. Joyce's two conditionals, $M \supset [(A \Box\rightarrow M) \& (B \Box\rightarrow M)]$ and $\neg M \supset [(A \Box\rightarrow \neg M) \& (B \Box\rightarrow \neg M)]$, come out as true or highly likely because M (or $\neg M$) is an effect of something in the causal history of A and B , and so gets held fixed when we evaluate counterfactuals with A or B or their negations in the antecedent. One need not follow Lewis in taking these non-backtracking counterfactuals to provide an analysis of causal claims to recognise that they are intimately connected with the idea of singular causation. Part of that idea is that causes explain effects because they make a difference. Singular causal claims thereby support a

range of modal judgments as to how the subsequent history would have been different had the cause occurred differently, or not at all.

But now the evidentialist may go on the offensive and press that the causalist's use of counterfactuals is objectionably unconcerned with the real world. Because the past history is held fixed in their evaluation, the use of causal counterfactuals means that, in general, certain outcomes—pairings of actions and states of the world—will be treated as live possibilities which are nevertheless not serious contenders as things that might actually happen. For instance, the outcome of taking both boxes and finding money in both. The reason for this is that the causal counterfactual is evaluated in possible worlds that deviate from the actual world in certain respects, perhaps involving small violations of natural law, or exceptions to known regularities. Some of these worlds will thus be highly improbable—for example, worlds in which one takes both boxes and there is money in the opaque one.³⁰ But deliberating agents are surely concerned with what their *actual* future will be like—not what happens in 'nearby' worlds that, for all their similarity to the actual world, are nevertheless highly unlikely.

Consider the first of Joyce's conditionals, "If the \$1,000,000 is in the bank, then it would still be there were I to do *A* and it would still be there were I to do *B*." This is something the agent should have a high credence in at any point in the problem, before or after making a decision. On the other hand, the only way they can *detach* the consequent pair of counterfactuals, "If I were to take the transparent box there would be £1,000,000 in the opaque box and I would receive £1,000,100", and "If I were to leave the transparent box there would be £1,000,000 in the opaque box and I would receive £1,000,000", is by deciding to leave the transparent box. Thus, the agent is only in a position to affirm the first counterfactual once they are confident that its antecedent is false. Revising their decision—if it is not too late—provides a basis for affirming the counterfactual's antecedent, but at the same time undercuts the basis for affirming the counterfactual. In general, Newcomb problems are characterised by the presence of causal conditionals such that the agent cannot consistently affirm both the conditional and its antecedent. It is then a pressing question for the causalist why the agent should regard such counterfactuals as action-guiding, since they can never give any reason to

³⁰Supposing, for the moment, that the predictor is near-infallible.

believe that the desirable state of affairs mentioned in the consequent will actually come about.³¹

This discussion brings out in more detail the tension between the use of causal thinking in explanatory and deliberative contexts. Causal explanation enjoins us to consider counterfactual worlds, as a measure of the difference the cause made to the effect. But it is not clear what bearing, if any, this kind of otherworldly evaluation has on situation of a deliberating agent. Insofar as one is concerned about the actual future, it just seems a mistake to act for the sake of outcomes that one is confident will not happen.

This tension might be cast as a conflict between ‘subjective’ and ‘objective’ takes on the agent’s situation. The causal counterfactuals undoubtedly reveal something about the choice the agent faced. There is a clear sense in which it is just true to say that that the one-boxer, despite ending up with a million, could have done even better had they taken the other box as well. The fact that their doing better was highly improbable, given what turned out to be the actual causal history of their choice, does not deprive the counterfactual of its relevance or interest. It may often be of interest to know how things would have gone differently had one acted otherwise, even if one’s acting otherwise was highly improbable given the other relevant causal facts. We should be able to accommodate these judgments as consequences of ordinary causal thinking; any analysis of causal notions which fails to do so is missing something out.

Nevertheless, there is a sense in which these judgments belong to a conception of the situation from outside the agent’s point of view. In the nature of the case, the agent has no access to the facts about the causal structure of their situation otherwise than by reaching a decision—by which point it is too late to exploit the causal conditionals in action. On this basis the evidentialist can reasonably deny that the one-boxer and the two-boxer do face different situations. Conceived from the agent’s point of view, they face exactly the same options: taking one box and likely getting a million, or taking both and likely getting just £100.³²

³¹Edgington 2011 makes much the same point. There are counterfactual conditionals which get a high probability only because the agent’s credence is distributed mostly across worlds where their antecedent is false—‘and that gives us no reason to act on it.’ (p. 84).

³²Some decision theorists also claim that the agent’s perspective is distinguished from an observer’s by the epistemic standpoint from which the agent’s options are to be evaluated. Thus Price 1991 invokes ‘agent probabilities’ for assessing options; and Joyce 1999 writes: ‘[The difference between EDT and CDT]

The challenge for the causalist is therefore to explain how the counterfactuals that characterise the objective causal structure of a problem can be of relevance to an agent whose involvement in the problem prevents them from affirming those very counterfactuals. This is an instance of the more general challenge raised at the beginning of this section: how to combine an objective understanding of causality with an account of how causal structure can serve as a guide to rational deliberation. The following section will explore how this tension plays out in the context of an account of causal inference that accords a central role to the connotations causal structure has in agency and planning, while providing this role with a more objective core.

3.4 INTERVENTION AND DECISION

This approach takes as central the notion of an *intervention*. Informally, the idea of an intervention is that of an exogenous influence on a system, which ‘surgically’ affects some component of the system without influencing anything else except via the component it affects. The relationship between the notions of intervention and cause can be roughly stated as follows: X is a cause of Y just in case X and Y are correlated, and continue to be correlated when X is intervened on. The intuitive idea here is that the asymmetry of causal relations is revealed in the fact that causes are ‘handles’:³³ events can be manipulated by way of intervening on their causes, but not by intervening on their effects.³⁴

The interventionist approach can represent the causal structure of a system by means of *causal graphs*. A causal graph consists of nodes connected by directed arrows,

concerns the appropriate *epistemic state* from which value judgments are to be made. Evidential decision theorists tell us to evaluate acts by provisionally modifying our opinions in the way that best reflects our considered judgments about their evidential import, whereas causal decision theorists tell us to make this modification in a way that best reflects our judgments about their causal powers.’ (p. 180, original emphasis) By contrast, the distinction I am urging between perspectives is precisely the reverse. Price and Joyce agree that taking an agent’s perspective involves disregarding backtracking correlations. Whereas the point I am pressing is that this counterfactual mode of assessment is one that only becomes available once one has the kind of clear view of a situation that agents in Newcomb problems are, in the nature of the case, unable to take.

³³The description of causes as handles comes from Collingwood 1937.

³⁴The formal theory at the heart of the interventionist approach is pioneered in Pearl 2009; Spirtes et al. 2000. Woodward 2003 utilises the formalism to develop a philosophical account of causation. My exposition is principally due to Woodward.

where each node represents a variable whose range of values represents the range of possible properties that some aspect of the system can have. One can think of the variables as representing determinable features of a causal system. The directed arrows represent causal relationships between variables. For instance, the graph $X \rightarrow Y \rightarrow Z$ represents a causal structure in which X causes Y , and Y causes Z ; and the graph $Y \leftarrow X \rightarrow Y$ represents a structure in which X causes both Y and Z . (The notion of ‘cause’ here is highly general, so that X causing Y can mean either that X is a promoting or an inhibiting cause of Y .)

The causal graph conveys information about the statistical relationships between the variables represented by its nodes on the assumption that the graph satisfies the *Causal Markov Condition* (CMC):

CMC: For any variable X in the graph, X is probabilistically independent from all other variables other than its *descendants* (i.e. nodes connected to X by a path of arrows directed away from X) conditional on the values of its *parents* (i.e. nodes connected to X by a single arrow directed towards X .)³⁵

For instance, by CMC, if two variables X and Y have a common parent Z , then conditioning on the value of Z renders X and Y independent: $P(X, Y|Z = z) = P(X|Z = z)P(Y|Z = z)$. Given enough information, CMC can thus be used to distinguish between different causal structures, for instance between the causal chain structure $X \rightarrow Y \rightarrow Z$ and the common cause structure $Y \leftarrow X \rightarrow Z$. In the latter case, Y and Z should be independent if a specific value for X is conditioned on, whereas this is not implied for the former. On the assumption that all correlated variables are connected by a path of arrows representing causal relationships,³⁶ this amounts to the Reichenbachian idea that common causes ‘screen off’ correlations between events.

³⁵Since the notion of cause here is neutral between promoting and inhibiting causes, CMC rules out both positive and negative correlations between causally unconnected variables.

³⁶Strictly speaking, correlated events must be connected by either a directed path (that is, with all arrows pointing in the same direction) or a fork (with arrows leading out in both directions from a central node: e.g., $X \leftarrow C \rightarrow Z \rightarrow Y$ is a fork.) Variables connected by inverted forks or ‘colliders’ (e.g. $C_1 \rightarrow E \leftarrow C_2$) are not correlated. Informally, directed paths reflect causal evidence, forks reflect ‘backtracking’ evidence, and inverted forks are evidential dead ends: if you have information about one of the typical causes of some event, you cannot infer anything about its other causes.

Applying CMC to a graph thus allows us to infer patterns of conditional dependence and independence between variables, and so constrains which probability distributions are admissible over the variables represented in the graph. But the causal graph represents more than the statistical information entailed by CMC. This can be seen from the fact that the same statistical patterns may be consistent with multiple different causal graphs that satisfy CMC. To take a very simple example, the graphs $X \rightarrow Y \rightarrow Z$ and $Z \rightarrow Y \rightarrow X$ entail exactly the same constraints on the correlations between X , Y and Z , namely that X and Z are independent conditional on Y .

What the graph conveys, in addition to statistical relationships between variables, is how these relationships are affected by interventions on the system. We can capture this by introducing the notion of an *intervention variable*. A variable I is an intervention variable for X with respect to Y just in case:

- (i) I causally affects X .
- (ii) I cancels the causal influence of X 's normal causes (its parents in the graph) on X .
- (iii) I causes Y only through its influence on X .
- (iv) If any further variable Z is correlated with I and causes Y , then there is a directed path from X to Y through Z .³⁷

If we draw a new graph introducing the intervention variable I , then, on the assumption that all causal relations are represented by arrows in the graph, condition (ii) then amounts to saying that the intervention 'breaks' any other arrows leading into X , and (iii) to the requirement that any directed path from I to Y go through X . We can thus represent the effect of an intervention with the altered graph, with a single arrow from I to X and all other arrows into X removed. (Interventions can be thought of as 'graph surgery'.)

Applying CMC to the altered graph thus delivers a new set of conditional probabilities, the intervention probabilities. We can then give a sufficient condition for a X being a cause of Y using the notion of intervention:

C-Suff: If I is an intervention variable for X with respect to Y , and, applying CMC to the graph altered for I , X and Y are not independent conditional on I , then X causes

³⁷See Woodward 2003, pp. 94–114 for discussion of these conditions and comparison with alternative formulations.

Y.³⁸

The difference between the original and the altered graph captures an intuitive distinction between two ways of discovering information about a system's causal structure: by passively observing it, or by acting on it. Information about how a system behaves when left undisturbed has a different significance from information about how it responds when interfered with in various ways. The latter kind of information has a depth that the former lacks; it reveals not just correlations, but relations of asymmetrical dependence, by interrupting some directions of influence while leaving others fixed.

There is good evidence that humans can learn about causal structure in this way. In experiments carried out by Alison Gopnik and colleagues,³⁹ test subjects (children and adults) were shown an apparatus (for instance, a box with three balls on sticks moving up and down) moving first with no apparent interference, and later as a result of visible manipulation by the experimenter (for example, the experimenter moving one of the sticks up and down by hand.) In some conditions subjects were also invited to experiment with the apparatus themselves. In general, subjects were sensitive to the difference between information about how the apparatus behaved when apparently left alone, and how it behaved when manipulated. Subjects were able to utilise this in drawing causal inferences, which they then drew upon both in planning their own interventions, and in answering explicit questions about the causal structure of the machine (for instance, which sticks caused the other sticks to move.)

One of the attractive aspects of the interventionist picture is thus that it fits very neatly with a plausible epistemics of causality. This is a sharp contrast with, say, Lewisian counterfactuals, with their complex and unwieldy inter-world similarity met-

³⁸The formulation stated here gives a sufficient condition for the notion of a *contributing cause*, represented by a directed path of any length. It is not a necessary condition because X could be a contributing cause of Y and nevertheless fail to be correlated under intervention, if X has additional effects that cancel out its causal influence on Y. Thus we also need a notion of *direct cause*, for which we can give a more complex, necessary and sufficient condition: X is a direct cause of Y just in case X and Y are correlated under an intervention which fixes this value of X, and an additional intervention which fixes the value of every variable other than X and Y. (Cf. Woodward, pp. 38–61.) However the distinction between direct and contributing causes is not important in what follows.

³⁹e.g. Gopnik and Kushnir 2003, 2005; Gopnik, Kushnir, Lucas et al. 2009; summary and discussion in Gopnik, Glymour et al. 2004; Gopnik, Kushnir and Schulz 2007. Gopnik et al in their 2004 also make the further claim that the formalism of causal graphs described above can model the underlying representational structures that enable this kind of learning. See also Hagmayer 2016; Sloman 2005.

ric. A key feature of the causal graphs formalism is that, in concert with the Causal Markov Condition, the graph acts as a kind of translation manual between information about the behaviour of a system under observation and under intervention. Predicting the value of some variable Y given an observed value of some other variable X corresponds to ordinary conditioning on the value of X ; whereas predicting the value Y given an intervention on X involves considering the altered graph, with the arrows into X broken, and conditioning on the value of X for the probability distribution entailed by CMC for the new graph. Being able to translate between these kinds of information is a central part of human causal thinking.

The interventionist approach thus captures nicely some of the connections between causation and agency: first, that causes are ‘handles’ or ‘effective strategies’, and can be used to bring about their effects; secondly, the corollary of this, that acting to intervene in one’s environment is a rich source of information about causal structure. In this respect, the interventionist approach has a certain amount in common with one like Price’s, which explains causality in terms of the structure of deliberation. It is worth commenting on some of the principal differences.

The first main point is that the interventionist approach does not offer anything like a reductive definition of cause, as the notion of cause is freely used in defining an intervention. This is a key contrast with agency-based approaches to causality. Accounts like Price’s explain the notion of cause by in terms of the role that a relationship between variables can play for a deliberating agent, where this role is characterisable fundamentally in evidential, non-causal terms.⁴⁰ This does not render the interventionist conception of cause as partially stated in condition C-Suff trivial, though, as the condition that X causes Y does not occur in the sufficient condition it provides. C-Suff states an informative constraint on how different aspects of causal structure must be related.

Secondly, the notion of an intervention is basically a technical one. The term ‘intervention’ suggests human agency, and this may be a helpful way of getting a handle on the intuitive idea, but this is not the level at which the notion is fundamentally explicated. Being a voluntary action is neither necessary nor sufficient for being an intervention.

⁴⁰As Price 2017 notes in his response to Woodward, this does not commit him to characterising agency itself in non-causal terms. The point is that the rational deliberative role of causal information can be characterised without appeal to its being causal.

Having said that, the technical notion of intervention does plausibly capture a property of actions that we often have reason to be interested in. Especially in controlled contexts, like scientific experiments or medical procedures, one may strive to ensure that one's actions come as close as possible to being interventions with respect to the properties of the system one is interested in, and one may want to know in what respects one's actions fail to be interventions in that sense.

A subsidiary point is that being an intervention on one variable is always relative to some other variable or variables. The above definition was of an intervention variable for X with respect to some other variable Y . One might wish to introduce a notion of an intervention on X with respect to the entire system of interest. But this notion is still relative to the set of variables under consideration. If we expand the system to include further exogenous influences on its outer nodes, events that previously were considered interventions may no longer be so. At the limit, if we consider the entire universe as a causal system, no event can count as an intervention.

There is a consequent sense in which causality is a local phenomenon. As Pearl (2009, pp. 419–420) puts it, 'If you wish to include the entire universe in the model, causality appears because interventions disappear—the manipulator and the manipulated lose their distinction.' However this does not imply—and Pearl surely does not take it to imply—that the facts about what causes what are relative to a model, so that there is no causation when the whole universe is taken into account. As stated by C-Suff, for X to be a cause of Y requires just that they be correlated under any intervention on X with respect to Y . The sense in which causality 'disappears' is just that causal relationships are not *discoverable* when one attends solely to the global pattern of events. At this level of description, the only regularities which appear are those described by time-symmetric dynamical laws. Discerning causal structure requires isolating subsystems of the universal order in order to see how those systems behave under exogenous influences. But none of this should be taken to imply that causal relationships are any less real, or that a complete description of the universe from no point of view would not include causality.

The interventionist approach thus makes causality a more objective matter than an agency-based approach does, even if the precise sense in which this is so may be some-

what difficult to pin down.⁴¹ Interventionist causality makes contact with the interests of human agency, but it does so by identifying something that we care about in deliberation: namely, dependencies between environmental variables that are preserved when an external event affects some variable otherwise than through the normal pathways of causal influence, while leaving other variables unaffected. And although this feature has an intimate connection with agency, it is not defined in terms of its role in agency as such, but rather in non-reductive, causal terms.

Given that the relevant notion of an intervention is thus an idealised, technical notion, captured in the conditions (i)–(iv) above, one might ask how it relates to real-world causal reasoning. The formal notion plausibly captures something of interest and value to agents and causal reasoners. But how is this best understood?

Part of an answer is that, as a matter of fact, our actions frequently do qualify as interventions with respect to the system on which we are acting, and we are warranted in treating them as such. In working out how to operate a new radio alarm clock, for example, one obviously does not need to consider the possibility that there is a hidden common cause of one's pressing a certain button and the snooze function being activated. Moreover this assumption is clearly operative in the causal learning behaviour described above. It is because the experimenters' or the test subject' manipulations of the apparatus were presumed to be exogenous influences that the different behaviour of the apparatus when manipulated could be taken as evidence of causal structure, and so relevant to answering questions like 'Which stick makes the other ones move?'

This raises the question how one should think about cases in which actions are not interventions with respect to the system in question. This is pointedly the case in Newcomb problems, where it is part of the definition of the problem that one's decision and the contents of the boxes have a common cause. But in more realistic cases, it may be that one has no way of knowing whether one's action and something one cares about have a common cause.⁴²

⁴¹For discussion of this question, see Woodward's (op. cit., pp. 123–127) critical discussion of Price and Menzies 1993; and Price 2017's response.

⁴²Should one take the decision to smoke in a medical Newcomb problem to be an intervention? On the face of it not, since smoking is correlated with the gene. Note, though, that the most straightforward way in which the 'tickle defense' is applicable is if the gene is presumed to operate by causing a first-order inclination to smoke, which may or may not be endorsed on careful deliberation. In this case, it may be

On the other hand, there is an equally good sense in which, in considering one's action retrospectively, one does consider it as an intervention. For instance, in the Newcomb problem, if the one-boxer looks back and says 'If only I had taken both boxes, I would be £100 richer', they are considering a counterfactual scenario in which the connection between their decision and its normal causes is broken.

What this brings out is that there are two senses of 'considering one's action as an intervention', or more generally two ways in which one can treat the hypothesis that some event e is an intervention. One can suppose that the actual causal structure is such that e meets the conditions of intervention; that e is an event $I = i$ of an intervention variable taking a specific value. Or alternatively, one can consider counterfactual scenarios in which e occurs without its normal causes.⁴³

Champions of the interventionist approach to causality note that the difference between observation and intervention, understood as a difference between ordinary conditioning on the value of a variable and conditioning on its value under intervention, is closely parallel to the distinction between indicative conditionals and (non-backtracking) counterfactuals (e.g. Hitchcock 2001; Meek and Glymour 1994; Woodward 2007.) If we condition on the value of some variable, via CMC, we get a new distribution of probabilities for both its descendants and its ancestors; in other words, effects are probabilistically relevant to their causes as well as vice versa. By contrast, if we intervene on a variable, we fix its value independently of its typical causes, so that conditioning on the new value leaves the probability distribution over its ancestors unaffected. (In other words, applying CMC to the altered graph entails that the intervened variable is probabilistically independent of its typical causes.) Thus conditioning under intervention is effectively conditional reasoning with a ban on 'backtracking' from causes to effects.

However this point requires some care. While it is true that these two distinctions—

right that the gene is not a cause of one's deliberation, and so (barring further hidden common causes) one's decision is an intervention. If, on the other hand, the gene causes you to smoke by influencing your deliberation—e.g. by causing you to adopt a causalist decision rule—then the decision cannot be an intervention. Meek and Glymour 1994 discuss the psychological conditions required for one's action to be an intervention.

⁴³This might mean considering e in possible worlds in which a Lewisian small miracle occurs. But this is not the only way of evaluating the counterfactual: the causal graph formalism can be used to provide an interpretation of the intervention counterfactuals, e.g. Halpern and Pearl 2005.

between conditioning and intervening, and between indicative and counterfactual conditionals—are analogous, they are not the same distinction. The notion of an intervention is not in itself essentially modal or counterfactual. An intervention is a type of event, an intervention variable's receiving a certain value, conditioning on which gives a new probability distribution for the intervened variable without affecting the values of its normal causes. When considering a hypothetical would-be intervention on a system, we can consider it in two different ways, just like any other hypothetical event: as actual or as counterfactual. That is, we can embed the hypothetical event in an indicative or a counterfactual conditional.

Often the distinction between an actual and a counterfactual intervention is unimportant, because the two notions coincide in terms of their consequences for the distribution of probable values over the system. An intervention variable is, by definition, not correlated with any variables in the system other than its effects; so, if an event is a genuine intervention, there should be no backtracking consequences of the conditional “if *I* occurs...” that are excluded from the counterfactual “if *I* had occurred...”. But, of course, this holds only as long the consequent of the conditional concerns only values of variables within the system intervened upon. As soon as we consider conditionals whose consequents concern exogenous conditions under which the intervention took place, there may well be a difference between backtracking and non-backtracking interpretations. Moreover, it may be that many intervention-like events—for instance, human actions—are only imperfect instances of interventions, if there are small correlations between their occurrence and other aspects of the system intervened on. In such cases, there will be a difference between considering the ‘intervention’ as it occurs in the actual world, and in idealised non-actual circumstances where it meets the full conditions of intervention.

This point is brought out by these remarks of Woodward:

...the notion of an intervention [as officially characterised]...represents a regulative ideal. Its function is to characterize the notion of an ideal experimental manipulation and in this way to give us a purchase on what we mean or are trying to establish when we claim that *X* causes *Y*...for this purpose, it isn't necessary that an intervention actually be carried out on *X*. All that is required is that we have some sort of basis for assessing the truth of claims about what would happen if an

intervention *were* carried out. (2003, p. 130)

Woodward's point is that, given the messiness and complexity of the causal structures that exist in the actual world, it may be that many actual intervention-like events fail strictly to satisfy the conditions of intervention: even in the most carefully designed experiment, there may for instance be collateral effects which slightly disrupt the causal relations under investigation. Nevertheless, our assessment of causal claims is not hostage to the possibility of devising a perfect experiment that can actually be carried out. Rather, the notion of an intervention can serve as a guide—in particular, as a guide for constructing and evaluating counterfactual scenarios in which we alter the cause while holding the rest of the structure fixed. This is a far more principled and practicable way of assessing causal counterfactuals than, say, Lewis's inter-world similarity metric.⁴⁴ But it brings out sharply the difference between considering, counterfactually, an ideal intervention for the purpose of evaluating causal claims; and considering, as an actual possibility, an intervention-like event in order to determine what its occurrence would tell us about the actual world.

The point is that, in certain contexts, there may be value in considering one's action in a counterfactual scenario where it is an intervention with respect to the variables of interest—even if this is not the causal structure in the actual world. This is quite different from making causal inferences about what will actually happen on the basis of the assumption that one's action is indeed an intervention. In a Newcomb problem, the latter procedure is clearly unjustified. On the other hand, we might construe the disagreement between the evidentialist and the causalist as precisely a difference concerning the legitimacy of the former procedure in the context of deliberation.

Nevertheless, the question remains why one should treat one's action as an intervention, in any sense, when in fact it is not. The problem is that the intuitive idea of causes as handles makes sense originally in the context of a situation in which one is acting as an exogenous influence on one's external environment, like shaking a tree to make fruit fall off. When, on the other hand, one acts on a system in which one is oneself enmeshed, this intuitive idea of cause breaks down and it is no longer clear what one should do.

⁴⁴An additional advantage the interventionist interpretation of counterfactuals claims over Lewis's approach is that it easily produces interpretations of counterfactuals that track the causal facts in cases of overdetermination and pre-emption. Cf. Hitchcock 2001.

This situation is dramatised in an especially pointed way in Newcomb's problem; but any reflective agent knows that this is in fact our predicament all the time, that none of our actions are interventions with respect to everything that we care about. When one considers oneself as a cog in the great causal machine, the only way to preserve the intuitive idea of causes as ways of manipulating their effects is by going 'off-world', considering one's action in counterfactual scenarios where it occurs without its usual causes. But it is hard to see how this procedure could be justified just on the basis of the idea that causes are effective strategies. Strategies that secure the desired goal only in non-actual worlds are hardly effective.

3.5 RETROSPECTIVE CAUSALITY

I have been developing the idea that there are two faces to the notion of cause, one connected with deliberation and manipulation, and one connected with explanation. I have also suggested that there is a sense in which the deliberative use is agent-centred, whereas the explanatory dimension captures something of the objective core of the notion of cause. The difficulty in reconciling them comes when one reflects on the causal structure of one's own agency—as one is forced to do in Newcomb problems. Here there threatens to be a rift between the agent-centred pursuit of one's goals, and the objective understanding of the causal structure of one's situation expressed with the explanatory use.

What I want to suggest is that there is a particular kind of explanatory use of causal concepts, and of correlative counterfactual conditionals, that is tied to a retrospective perspective on events. One might call this specific mode of causal thought *historical* causal thinking. In using causal concepts in this way, one explains a certain actual course of events with reference to prior causes. As noted in section 3.2, at the ideal limit of this process all events are explained with reference to the initial conditions of the universe. But, as the previous section brought out, it is not obvious how this use of causal concepts is connected with their use in deliberation.

The two-boxer's best argument for their choice, I have argued, makes implicit reference to causal-historical explanation from a later point in time: after the prizes are

awarded, they can say to the one-boxer, “Given the situation you actually faced, if you had taken the transparent box too, you would have ended up richer.” In other words, they can argue that the one-boxer ought to regret their choice. Moreover, due to the Dominance structure of the payoffs in a Newcomb problem, the one-boxer, at the time of choosing, is in a position to anticipate that they will regret their choice *however* things turn out.⁴⁵

Of course the one-boxer may be more moved by the backtracking counterfactual reasoning: “If I had taken the other box as well, I would have been a different kind of agent, and been penalised for it.” And, if their choice was accurately predicted, they may find their reward ample compensation for any regret they might feel at the thought that, had they chosen differently, they could have got even more. But the two-boxer’s counterfactual expresses something that should still be of interest to any agent, whatever decision rule they use: it reveals how their own agency made a difference to the course of events. This is a deployment of causal thinking in connection with agency that is *prima facie* quite different from, and not obviously reducible to, the idea that manipulating causes is a way of manipulating effects.

One way of bringing out what is distinctive about this kind of causal thinking is to point out that causal judgments of this kind are unaffected by overdetermination and pre-emption. If Suzy throws a rock which smashes a window, the judgment that Suzy broke the window is not undermined by the fact that Billy’s rock, launched a second later, would have broken it had Suzy’s missed. By contrast, when we consider the perspective of pure deliberation, pre-emption and the like *does* undermine the significance of causal connections. The fact that a certain event E will come about anyway regardless of whether one does A undermines one’s reason for doing A for the sake of E; indeed, if doing A is somewhat costly, then doing A is strictly dominated by not-A, and so irrational by the lights of standard decision theory (evidential or causal.) Thus, if A is

⁴⁵This is the crucial respect in which Newcomb differs from problems like Death in Damsacus, where you know that Death waits for you in either Baghdad or Damsacus and has predicted where you will go. Here each choice is (non-causally) correlated with a state of the world that makes the other choice the better option, and so one can know in advance that one will regret whatever one chooses, without knowing which specific choice one will regret. Thus the method of anticipated regret just described cannot yield a stable recommendation. Egan 2007 describes a family of problems with essentially the same structure. For discussion of the instability exhibited by these cases in relation to CDT, see Arntzenius 2008; Joyce 2012.

pre-empted as a cause of E, A is not a way for an ideally rational agent to bring about E.⁴⁶ One question this might prompt is whether deliberative or agency-based accounts of causation can come up with the right judgments in cases of overdetermination and pre-emption. Perhaps they can.⁴⁷ The more immediate point is just to draw attention to a way we employ causal concepts in connection with agency, in retrospectively attributing responsibility, that does not straightforwardly track the facts about how a fully rational agent would have deliberated at the time.

My proposal is that, in Newcomb and Newcomb-like scenarios, justification of causalist reasoning must ultimately rest on an appeal to an anticipated retrospective perspective on the problem, one from which causal concepts and associated counterfactuals are used to track agential responsibility. Without bringing in this perspective, the causalist does not have a satisfying answer to the evidentialist challenge discussed in section 3.3: they can provide no reason to care about what goes on in ‘nearby’ but improbable counterfactual worlds.

The idea that knowledge of one’s future informational state can constrain how one should reason at an earlier time is not unfamiliar: it is embodied in van Fraassen’s (1995) ‘reflection principle’; and in Jeffrey’s (1981) ‘ratificationist’ amendment of EDT to deal with Newcomb-like cases. The present point is not about the specific form that a theory of rational belief or decision should take, however. The point is rather about the external justification of causalist reasoning, that it involves a use of causal notions which goes beyond the idea that causes are ways of bringing about effects. This is consistent with a range of ways in which broadly causalist reasoning might be implemented in the context of a formal theory.

This chapter has been concerned with a general dilemma that arises for the idea that the difference between past and future can be understood purely in terms of the structure of deliberation. On the one hand, from the internal point of view of deliberation itself there does not appear to be anything intrinsically absurd about acting to change the past. On the other hand, the best candidate for an objective temporal asymmetry

⁴⁶This point is made by Hitchcock 2013, 2017.

⁴⁷For instance, Fernandes 2017’s biconditional related causation and deliberation—that C causes E iff a rational agent deciding on C for the sake of E for proper deliberation would be evidence for E—is not obviously subject to counterexamples involving pre-emption and overdetermination.

in the world that might explain or justify a matching asymmetry in deliberation is the structure of causality. But, in the cases where the asymmetry of deliberation comes under greatest pressure, namely Newcomb problems, it is not obvious why the direction of deliberation should follow that of cause to effect.

The suggestion raised in this final section is that we should recognise a species of causal thinking that bears on agency otherwise than through deliberation. Judgments of agential responsibility are causal in a way that is not directly answerable to the dependencies that could have been exploited by a deliberating agent. Moreover, if this use of causal concepts is essentially retrospective, then there is an important sense in which the temporal perspective of agency is not just the perspective of deliberation. Rather, the relevant feature of an agent's perspective is that there is a non-deliberative way of thinking about one's own agency which can only be had retrospectively. This retrospective stance is the topic of the next chapter.

4 | REGRET AND RETROSPECTION

The previous chapter argued that the significance of causal thinking cannot be explained purely in terms of its bearing on deliberation. This was posed primarily as a question about the justification of causalist reasoning in Newcomb and Newcomb-esque decision situations. The key point was that such reasoning cannot be justified by appeal to the idea that causes are ways of bringing about their effects, but effects are not ways of bringing about their causes. In a Newcomb problem, precisely what is at issue is whether an effect might be used as a way of bringing about its cause. This point was not, however, intended as a refutation of causalism, but rather to draw attention to a dimension of causal thinking as relating to something other than agentially exploitable dependencies.

The more general idea is that there is an area of causal thinking—thinking about the causality of one’s own past agency—which does more than merely trace the contours of information that was available for deliberation, even to an ideal deliberator. Retrospective causal understanding comes apart from deliberation in this way when there are aspects of causal structure in principle unavailable from the standpoint of deliberation. In the highly artificial and sterile environment of a Newcomb problem, this feature is ensured by stipulation: there is no way for the agent to obtain evidence about the Predictor’s choice otherwise than by deciding. One might therefore take Newcomb problems to be no more than a theoretical limit case in which the equation of causal structure with agentially exploitable relationships breaks down. As I shall argue, however, when we take into account the complexity and messiness of human motivation, structural limits on what one can know *qua* deliberator about the causal structure of one’s own agency are in fact ubiquitous.

The aim of this chapter is twofold. The first is to point towards, and to clarify, a kind

of concern we have for our own past actions. The second is to argue that this concern presupposes a species of understanding that one can have only towards actions which are completed. This existence of this distinctively retrospective perspective implies that the temporal perspective of an agent is not to be understood just in terms of the structure of rational deliberation. Rather, on the picture that will emerge, deliberation is a kind of thin superstructure that floats on top of, rather than explaining, the deep structure of agency in time.

In section 4.1 I draw on Bernard Williams's discussion of agent-regret to motivate the idea of a kind of retrospective assessment that relates distinctively to one's own actions. In sections 4.2 and 4.3, I put this intuitive idea on a firmer footing by arguing that, very generally, one's thoughts about one's own actions are subject to temporal constraints in virtue of the processive and dynamic character of motivation and action. I take as a paradigm case the attitude of relief that an unpleasant experience is over, arguing that the temporal profile of relief must be understood in terms of its relation to the motivational process of aversion. I then argue that intentional action, and the knowledge one has of one's own actions, is fundamentally a similar kind of motivational process, and that this places more general temporal constraints on the kinds of thoughts one can have about one's action while one is acting. I contrast this with the model of motivation that emerges from considering agents as fundamentally maximisers of expected utility. I close in 4.4 with some reflections on the practical value of relief, and its relation to a conception of oneself as extended in time.

4.1 AGENT-REGRET, DISAPPOINTMENT, AND REMORSE.

In 'Moral Luck', Bernard Williams aims to identify a distinctive kind of retrospective attitude he calls 'agent-regret'. He distinguishes it in this way:

The constitutive thought of regret in general is something like 'how much better if it had been otherwise', and the feeling can in principle apply to anything of which one can form some conception of how it might have been otherwise...In this general sense of regret, what are regretted are states of affairs, and they can be regretted, in principle, by anyone who knows of them. But there is a particularly important species of regret, which I shall call 'agent-regret', which a person can feel only towards his own past actions...In this case, the supposed possible difference is

on that possibility, the thought being formed in part by first-personal conceptions of how one might have acted otherwise. (1981, p. 27)

The distinction Williams makes here is intuitive enough. The OED has two separate entries for (non-obsolete uses of) the word ‘regret’: the first, ‘Sorrow, distress or disappointment due to some external circumstance or event’, does not imply any intrinsic connection with the past; and a number of the examples given are not concerned with past events. (The example given is: ‘Several residents to-day expressed their regret at the probable destruction of this picturesque piece of country.’) The second entry, on the other hand, does make the connection with the past, and also, notably, with agency: ‘Sorrow, remorse, or repentance due to reflection on something one has done or omitted to do.’ This suggests the thought that we have basically two notions of regret: a more schematic one, understood as a general negative reaction to the world; and a special one, whose application is restricted to past exercises of one’s own agency.

In its most schematic form, the first, general notion of regret just comes to: one has some preference about how the world should be, and believes that things are not that way. Presumably this general attitude can be differentiated by many subtleties and shades of affect, perhaps associated with kinds of subject-matter, degrees of intensity, connections with motivation and action, and so on. Without regard for any of these differences, let me call this highly general negative attitude *disappointment*. Anyone capable of forming evaluative attitudes of any kind is liable to be disappointed by the world when it fails to live up to their hopes and expectations. One can be disappointed that England are unlikely to regain the Ashes in 2019; that half the world’s wealth is owned by a few billionaires; or that no finite and consistent axiomatisation of arithmetic is complete.

How is agent-regret marked off from the more general attitude of disappointment? One thought is that agent-regret is just disappointment with a special subject-matter, namely one’s own actions. But Williams, at least, has a stronger characterisation in mind: he writes that the counterfactual alternatives involved in agent-regret are ‘formed by first-personal conceptions of how one might have acted otherwise’, rather than merely how things might have been otherwise.¹ Let us see if we can make any more of this.

¹Williams does offer a further criterion for distinguishing agent-regret from general regret. This is

There is a decision theorist's notion of regret, defined as the difference between the utility of the best act given in actual state of the world, and the utility of the chosen act given in actual state. Regret has a non-zero value whenever the agent's act was not in fact the optimal one given the actual state of the world. The presence of regret in this sense is consistent with one having chosen well, given the available information: for instance if the chosen act had the highest subjectively expected utility among feasible alternatives given one's credences at the time of choosing. Suppose you are offered a choice between three bets on the outcome of two fair coin tosses: £1 if both heads, £1 if both tails, £1 if one heads, one tails. Being rational, you bet on one heads, one tail; unluckily, they both land heads and you win nothing. Your 'regret' in this situation is equal to the utility of £1. But this carries no implication that you chose badly.

It is hard to see how the decision theorist's notion could be of much interest to individuals' evaluations of their past actions.² It registers that, had one chosen differently, things would have been better, but in doing so it does not purport to identify any kind of fault in one's conduct. There is only a very minimal sense in which it is centred on one's agency at all. It is true that it selects counterfactual scenarios by varying the chosen act and holding fixed the actual state; but when one is confident that one chose rationally this counterfactual assessment may have very little force. In the coin toss example, for instance, it seems to get things wrong to regret betting as one did, when it was so clearly the appropriate thing to do. Rather, the suboptimal result is due to unfortunate external events not under one's control. In these circumstances it seems unreasonable, if not downright irrational, to regret one's choice.

One way of making out a distinctive form of regret for one's own actions is to tie it to a sense of one's own conduct failing to meet some normative standard which one might

the idea that, if one sees oneself as agent of the regretted occurrence, one may feel motivated to some kind of redressive action that goes beyond compensating for any material harm or loss. (Ibid., p. 28) This criterion can hardly be sufficient, though, because, as Williams acknowledges, in many cases 'there is no room for any appropriate action at all.' (p. 29) Moreover focusing on Williams's criterion of redressive action is liable to bring in further complications concerning apology, forgiveness, and so on; I will therefore largely set it aside.

²Indeed Savage 1951 proposes that his 'Minimax regret' principle—the principle of choosing the option with the lowest possible regret value—is most readily applicable in group choice situations: 'Application of the minimax principle from this point of view means to act so that the greatest violence done to anyone's opinion shall be as small as possible.' (p. 62)

reasonably have been expected to meet. This form of regret is sometimes distinguished as *remorse*, and thought to be especially connected with blame and moral responsibility. Marcia Baron distinguishes: 'Remorse involves not just a judgment that it would have been better if one had acted otherwise...It involves the judgment that one could have acted otherwise, and, more important, that one *should* have.' (1988, p. 267) We might add that the force of the 'could' here is importantly stronger than a 'might': it is not just that there was a time when, for all that could be determined, one might have been going to act otherwise from how one actually did. The 'could' is rather bound up with the normative 'should': it means something like that there were at the time reasons against acting as one did of which one was aware, and it would have been reasonable to expect one to be moved by those reasons.

The exact content of these modal judgments is admittedly obscure, and what I have to say by way of elucidation is only partial and provisional. The relevant point is that remorse traffics with a conception of responsibility, connected with the kind of control an agent exercises over their actions through their sensitivity to reasons, which is stronger than the basic notion that a certain change is attributable to a person's agency. In practice this means that judgments of moral responsibility, and related reactions of blame and remorse, are modulated by considerations of ignorance, compulsion, mental disturbance, and so on; considerations which do not at all affect the judgment that a person was active rather than passive with respect to the occurrence in question. These qualifications might be summarised by saying that remorse is restricted to *voluntary* actions, or aspects of actions—although the notion of the voluntary is not necessarily any less obscure than the modal notions just mentioned.

Discussion of remorse tends to focus on actions which are contrary to morality; in the case of actions which are irrational it may perhaps be more appropriate to talk of self-reproach. Riding roughshod over any differences, the same broad dynamic may be supposed to apply when one reproaches oneself for acting irrationality as when feels remorse for wronging someone. The idea is still that there were considerations that spoke against acting as one did, which one could and should have been sensitive to at the time. Both morality and rationality purport to be, so to speak, a guide to life: a general-purpose code that can be used to guide one's conduct regardless of one's beliefs,

desires or external circumstances, such that whether one manages to abide by it or not is strictly one's own affair.³ Since a code of conduct of this kind applies exclusively to one's voluntary actions, the restriction of remorse to violations of this sort implies a restriction to voluntary exercises of one's agency.

This very broad characterisation of the role of morality and rationality would explain why remorse is an attitude one can have only to one's own actions. Although others may criticise or blame you, for failing to meet the demands of morality or rationality, there is a special force to this thought when applied to one's own case, connected with the fact that one is alone responsible for whether one complies with the demands of morality and rationality. Minimally, this means that the judgment that failed to do so ought to issue in a resolution to do better in the future. This conception likewise makes sense of why remorse should be essentially backward-looking. Adopting this same stance in advance of the act in question should similarly issue in a resolution to refrain from it. Except in unusual cases, such a resolution typically undermines one's belief that one will perform the act in question.⁴

Nevertheless the model of remorse is emphatically not what Williams has in mind with his notion agent-regret. He is explicit that agent-regret is not necessarily restricted to voluntary actions and their intended effects: 'It can extend far beyond what one intentionally did to almost anything for which one was causally responsibility in virtue of something one intentionally did.' (Ibid., p. 28) Here he gives the much-discussed example of the lorry driver who runs over a child despite driving safely and taking all reasonable precautions. Even though the driver is in no way at fault for the accident, he is unlikely to rest content with that judgment and view it as just an unfortunate external circumstance; indeed we should be somewhat suspicious of someone who reacted in that way.⁵

³This is a rather extreme characterisation: many people would argue that there are intrinsically irrational beliefs and intrinsically immoral desires, at the very least. Since my principal interest is in identifying a conception of agent-regret by contrast with the model of remorse, rather than a proper account of the nature of morality and rationality, I will stick with this caricature statement.

⁴In Phillip Pullman's young adult fantasy novel *The Amber Spyglass*, a killer priest does 'pre-emptive penance', and is absolved in advance, for an assassination he deployed is to carry out for the Church. The perversity of the idea of pre-emptive absolution is meant to be obvious.

⁵As a side note, the anecdotal evidence is that causing accidental death is a uniquely traumatising experience. Perhaps surprisingly, there is very little research into the psychological ef-

The suggestion that Williams's lorry driver exemplifies a distinctively agential form of regret is still controversial. R. Jay Wallace, in recent work, expresses scepticism that there is a distinctive way of relating to and evaluating one's own actions otherwise than through the structure of voluntary action and responsibility mapped out by remorse. He writes,

Perhaps...the exclusive character of agent-regret should be traced to something in the thoughts that it characteristically involves. But how exactly is this suggestion supposed to get fleshed out? The uninvolved observer or third party, after all, can perfectly well appreciate that the unfortunate event occurred, and also that it is unfortunate; the evaluative thoughts involved in these forms of awareness of the past event are precisely what make it possible for such third parties to experience impersonal regret about it. So what proprietary form of conscious awareness might be taken to be available only to the agent? Is it just the indexical thought that it was *my* agency that gave rise to the unfortunate event? It is true that an emotional reaction that constitutively involved an indexical thought of this kind would not even be available to those who were not themselves caught up as agents in the regrettable events...But it isn't obvious that emotions of retrospective assessment that involve such indexical thoughts really constitute an interesting natural kind of psychological phenomenon. (2013, p. 36)

Wallace does not mean to deny that one can be distinctively pained by the thought that one caused a certain unfortunate event. But, he argues, insofar as this special concern extends to products of one's agency that are non-voluntary, this should be seen as on a par with the attitudes one can take to misfortunes that befall one:

Our web-like history as persons...extends in two directions, involving non-voluntary causal influences on us as well as nonvoluntary causal effects that are brought about through our own agency. If this is the salient feature of our predicament, however, then there ought to be a form of regret that is adequate to both

fects of accidental killing or their effective treatment. A recent survey and study using qualitative interview methods is Rassool and Nel 2012. Some journalistic articles incorporating personal testimony include: <https://www.newyorker.com/magazine/2017/09/18/the-sorrow-and-the-shame-of-the-accidental-killer> and <https://www.theguardian.com/global/2018/nov/29/what-happens-to-your-life-after-you-accidentally-kill-someone>. A striking and persistent theme is the killers' unshakeable sense of personal responsibility for the deaths they caused, and the rupture in how they conceive their own agency in relation to the world. One interviewee in the New Yorker piece just cited expresses her frustration at the consolations of friends and family: "Yes, it was an accident, and in a certain sense we were both to blame, but, at the end of the day, I *hit* him, I took *his* life," she said. "No matter how much you want to dismiss it as an accident, I still feel responsible for it, and I am." She cried, "I hit him! Why does nobody understand this?"

of the ways in which we are linked causally to the larger network of events within which we operate. That is, not only should we expect people to feel a special kind of regret when untoward effects are inadvertently brought about as a result of their intentional actions. We should equally expect people to feel a special kind of regret when they in particular have been acted upon by the world in ways that are unfortunate. (Ibid., pp. 37–38)

Wallace's general thought is the following: granted, we can feel regretful about many products of our past agency. One source of such concern is when there is 'a lapse of agency on our part, a failure to do something that we ought at the time to have done.' (p. 40) But, as he says, this is 'the familiar domain of remorse, which attaches...to our voluntary performances.' (p. 41) On the other hand, in cases where it is really true that no blame attaches, that one was not at fault, the crucial factor is not that one was the agent of the occurrence, but simply that one was related to it in some way that gives one reason to care. The more fundamental relation he sees is just that of being partial with respect to some matter in virtue of one's personal involvement, where agency is just one way of so being involved. Correlative to this kind of partiality he introduces the notion of 'personal regret': an attitude which 'constitutively involves an appreciation of the special reasons that one's experience and attachments provide for feelings of pain and distress about something that happened.' (Ibid.)

The controversy over agent-regret relates to the present dialectic in the following way. I am arguing that there is a kind of causal understanding one has of one's own agency that comes only retrospectively. Causal judgments of this kind are not just judgments about the information that was available to one at the time of acting; they can include information that one may not have been in a position to affirm at the time. In the standard Newcomb case, the relevant information is, for example, that the £1,000,000 would still have been there had one taken both boxes. The question is whether this structure is limited to such cases, or whether something like it can arise in more mundane contexts.

The notion of agent-regret is suggestive of a wider array of distinctively retrospective attitudes to one's own agency, whose similarity to the causal knowledge one acquires on making a decision in a Newcomb problem we might explore. On the other hand, Wallace's treatment of agent-regret implies a sparser picture. On his picture, assessment of

one's role as agent is limited to how effectively one processed and responded to information that was in principle available at the time. Any insight one gains into the causality of one's past action is rather in the order of greater knowledge of the causal structure of the environment one was acting in—knowledge which, had one had it at the time, could have served as input to deliberation just like any other piece of causal knowledge. For example, one may think to oneself, "If only I'd known opening that door would set off the alarm, I would have gone the other way." But this piece of causal knowledge can hardly be thought of as in any way revealing of the causal structure of one's agency as such. It can seem that no more than this is involved in the case of Williams's lorry driver. He might think to himself "If I'd known she was behind that wall, I would have slowed down", or whatever. In general, the acquisition of some piece of causal information such that, had one known it earlier, one would have acted differently, does not seem to amount to a distinctively retrospective understanding of one's action in any interesting sense.

In order to get a grip on the kind of retrospective assessment I have in mind, and to bring out its connections with causal understanding, it may be helpful to have some examples in addition to Williams's lorry driver.

Suppose I make a remark about a colleague's work that comes across as rather disparaging. I offer it as a piece of friendly criticism (or so I think to myself), but a combination of tone of voice, conversational context, and so on, conspire to create an overall impression more negative than I intend. Thinking back later on the conversation, I realise that my colleague was put out. 'I shouldn't have said it like that', I think; 'I should choose my words more carefully'. Later still, I reflect that the tone of my remark may not have been so accidental: perhaps I was annoyed by an earlier remark of hers, without realising at the time; or perhaps it was the expression of an underlying aggression which I mostly prefer not to acknowledge.

I think that this pattern of ongoing self-evaluation and self-criticism, or variants on it, is familiar to many people. Events of this kind and our responses to them make up the much of the fabric of everyday social life. Yet it is not clear how to place these attitudes in a scheme that sharply divides responsibility proper from mere agential involvement.

It is clear that my remark and its conversational effects are not just an unfortunate

occurrence of which I happen to be the cause (indeed it is hard to describe what is unfortunate about it without making essential reference to my conduct—there is no impersonal description of any ‘harm’ available here.) But it also seems Procrustean at best to understand my fault here as of the kind that merits remorse, in the above sense of a failure to act on considerations that I was aware of and could and should have been moved by. It is certainly not the case that the disparaging effects of my remark were *voluntary*. Rather my failure, such it was, involved a lack of delicacy, a failure to pick up on relevant social and conversational cues and moderate my conduct accordingly. Even to the extent that, as I later suspect, there was a latent hostility behind my remark, this does not render it voluntary so much as evince a lack of self-control, and a failure to moderate my conduct in accordance with the motives of collegial respect that I normally endorse. Assimilating this to the case where the reasons and options were clearly before me elides the evaluative difference between a wilfully malicious comment and a merely crass, careless or intemperate one.

This is of course not to deny that I may have behaved obnoxiously, that my colleague has every right to be annoyed, I owe her an apology, and so on; insofar as this is the case, it registers the fact that our interpersonal relations involve a concern for character and dispositions beyond the narrower structures of evaluation and responsibility mapped out by remorse. But equally—and crucially—it may equally be that my colleague is not overly preoccupied with what I think of her, and that any subsequent attempt to apologise for my earlier remark would be a piece of pointless self-indulgence. This witnesses the equally important fact that the interest we take in our own actions often goes well beyond anyone else’s, and that self-criticism often attaches to features of our conduct that from an external point of view are not very significant.

What I want to stress is that the epistemic perspective from which I assess my remark, the perspective of retrospective self-criticism, is one from which I have an understanding of my action and the context in which it occurred that I could not have had at the time. Moreover, this is not just a matter of greater knowledge of what my action was likely to cause. (Indeed if my only reaction were “If I knew she would react like *that*, I wouldn’t have said it”, one may feel this betrayed a lack of appropriate sensitivity.) What is distinctive of the later perspective is not just knowledge of the effects of my action,

but an insight, or the possibility of an insight, into the state of mind that caused it. What allows for this insight is the possibility of adopting a reflective stance on my own agency and its psychological causes.

Elsewhere Williams offers an example of a different kind of retrospective assessment:

...consider the only too well-known case of a man who wants to be loyal to one woman and resolves to stop seeing another. He sees her all the same. If he finally goes back to the first woman he may classify these episodes as akratic, i.e. as cases of his doing other than what at the time he thought he had most reason to do. But if he finally leaves the first woman to live with the second, he may not see the episodes in this light, but rather identify them as the first steps in the recognition of what he had most reason to do. It was not a psychological defect that he could not achieve that recognition in full synchronous consciousness; perhaps it could not be achieved in any way other than that in which it actually came about. (1995c, p. 24)

In this kind of case, the retrospective assessment takes on an additional complexity because of its importance for the way the agent conceives of their life and what matters in it. One might call this aspect of retrospective thought its autobiographical, or narrative, dimension.⁶ To this extent the case may be taken to illustrate the ways in which retrospective assessment are informed by one's present attachments and values.⁷ The cynical version of this thought is that narratives about one's past actions may be recruited or constructed to bolster or reinforce one's current self-conception, rather than reflecting the truth about what happened. No doubt this is often the case. But this is all consistent with insisting on the point that what is at stake is, at bottom, causal information about the psychological sources of one's actions. The man in Williams's example is forming and evaluating causal hypotheses about his past actions, in a way that he was not able to at the time, when he was in the grip of the complex of psychological forces that later tries

⁶The use of the notion of narrative in connection with the explanation of life and action is closely associated with the work of MacIntyre 1981 and C. Taylor 1989. One may have reservations about excessive emphasis on the notion of narrative to the extent that it brings in interpretive concepts—like plot and genre—that seem of dubious relevance outside of literary contexts. This criticism is pressed by Williams 2009.

⁷This is the principal point Williams makes about the case. The situation is complicated by Williams's internalism about reasons, which raises the question whether the past-tensed phrase 'what one *had* most reason to do' is intended to capture the subjective reasons of the agent at the time of the retrospective assessment, or at the earlier time of the action in question. Some of the complexities arising from intertemporal shifts in agents' values and motivation are discussed in Williams 1995a.

to make sense of. The fact that such assessments may often be distorted or downright self-deceptive does not detract from their status as genuine causal judgments.

In general, it is by no means an unfamiliar thought that one often has insights into one's past state of mind, and the causes of one's past actions, that one was not in a position to have at the time. One might say to oneself, "I realise now that I only did it because I wanted so-and-so's approval"; or "I couldn't see it at the time but I was in a state of terrible anxiety that Christmas; that's why I acted so rudely to so-and-so." Equally familiar is the idea that there are certain things one can learn about oneself, about what kind of agent one is, only by acting.

Reflecting on these kinds of cases suggests we need a more nuanced picture of regret and its sources than emerges from Wallace's discussion. It does not seem to be the case in the above examples that one's role as agent gives rise to a distinctive way of relating to a past occurrence only when one's retrospective thought identifies some reason that one could and should have been sensitive to at the time. Rather, we should also acknowledge a way of relating to one's past action that flows from the kind of insight one gains only retrospectively. The rest of this chapter will attempt to put this idea on a firmer and more principled footing.

4.2 MOTIVATION AND PREFERENCE IN TIME

The above discussion focused on some examples in which it seemed intuitive to say that an agent gains some causal information about themselves after acting which they were not in a position to grasp at the time. But it might be thought that this does not point to any genuinely distinctive causal thinking that cannot be understood in terms of information available to an ideal deliberator. Such cases were rather characterised by the kind of lack of self-knowledge which marks out the agent as less than fully rational. According to this response, the *only* kind of case in which an ideal deliberator gains an insight, after acting, into the antecedents of their actions which they were unable to access before, is in a Newcomb problem, where this condition is built into the nature of the case.

Another way of putting this response is to say that ordinary causal thinking is gov-

erned by the regulative ideal of a fully rational deliberator with perfect introspective access to its own mental states. If we accept this regulative ideal, I think we should agree that it is in only in Newcomb-like cases involving unusual correlations between one's decision and the causal structure of the situation (predictors, twins, and so on) that there may be genuine mismatch between the agent's deliberative and retrospective perspectives. But the question is whether we should accept that regulative ideal. Clearly there are some emotional states—blind rage, for instance—such that being in it prevents one from knowing that one is in it. But perhaps an ideal deliberator would never be in such a state. In the following sections I will argue that such states are not just exceptions which can easily be set aside; rather, there limits on what one can know about one's own agency and when, in virtue of being an agent that is moved and acts in time.

Consider the desires that move us to action.⁸ Paradigmatic examples are simple animal desires, like those for food, sex and shelter. In the most basic case, these are expressed in behaviour aimed at bringing about their satisfaction. This motivational role of desire is encapsulated by Anscombe's slogan: 'The primitive sign of wanting is *trying to get*.' (1957, p. 68)

Considered as a source of motivation, desire is an occurrent, dynamic phenomenon, intimately connected with the causal-psychological processes of animate life. An agent is modified by the onset of some desiderative force—a drive, urge, or impulse—which moves it to engage in purposive activity aimed at satisfying some need. If and when the desired object is achieved, the satisfaction of the desire amounts to the cessation of its motivational force. The phases of these processes moreover involve characteristic changes in arousal and affect: the pursuit of desire involves high arousal, feelings of anticipation, and perhaps anxiety and distress if the attempt is thwarted; the attainment of the desired object is typically followed by low arousal and feelings of relief or satisfaction.⁹

This simple model of desire and motivation as a dynamic process may be expanded and etiolated in various ways. For a start, it is clearly a simplification to suppose that

⁸My discussion of the issues in the next few paragraphs is very much indebted to an unpublished manuscript of M. G. F. Martin (Martin Unpublished). Similar ideas also appear in Munoz-Dardé 2016 in connection with Wallace on regret.

⁹A more detailed account of the connection between desire and motivational processes along these lines is Wollheim 1974. Thompson 2008 also emphasises the processive nature of desire and intention.

the initiation of purposive action is always occasioned by the onset of a novel desire; it may also, for instance, be triggered by the perception of an opportunity to satisfy a standing, previously recessive desire. Moreover, sophisticated deliberators like humans are to some extent capable of ‘standing back’ from their desires in order to delay gratification, choose between incompatible courses of action, and so on. But allowing these many refinements does not undermine the central point that there are forms of wanting which are essentially dynamic, and express themselves through psychological processes that unfold in time. The occurrence of a desire in the most upstanding and self-controlled deliberator will still make itself felt to some extent in their affect and behaviour, even if they refrain from acting on it intentionally. Even if there is no change in someone’s observable actions, occurrent desire nevertheless affects conscious thoughts and feelings, and we might see this as a kind of inner shadow of the overt purposive activity that is the more basic expression of desire. (For instance: around 12 o’clock I get increasingly distracted from writing my paper by thoughts of what to have for lunch, although I am resolved not to break before finishing this section.)

Certain affective states are intrinsically tied to phases of the motivational cycle, and consequently have a particular intrinsic temporal orientation. This fact is exploited in A. N. Prior’s famous ‘thank goodness that’s over’ argument (1959). Prior’s argument turns on the fact that there is a certain emotion—relief that an unpleasant experience is over—which can be felt only at a certain time, namely, after the cessation of the experience in question. Prior takes this to be explained in terms of the relevant state of relief having a certain tensed content: the fact that the experience is over *now*—as opposed to the untensed fact that, e.g., it is over on April 25, 2019, which could in principle be appreciated at any point in time.

The fact that such emotions are tied up with dynamic processes puts Prior’s observation in a different context. The onset of an unpleasant experience prompts a desire for the experience to end; a desire which may be expressed in overt aversive behaviour, attempts to get away, and so on, or, if the agent suppresses these natural responses, felt purely privately. The cessation of the painful stimulus ends the aversive reaction, bringing feelings of relief. The state of relief in question is thus one which can be had only at the relevant point in the motivational cycle. This cycle of pain–aversion–relief is just

one instance of the kinds of dynamic psychological process connected with occurrent desire, and the place of relief in this process explains its temporal profile.

The above offers a very brief sketch of an aspect of the temporal perspective of an agent, specifically the perspective a relieved agent occupies on the occurrence at whose cessation they feel relief. We might pause to notice a few points about this kind of explanation.

First, the temporal orientation of the agent with respect to the occurrence in question is not explained just in terms of the agent's having a representation with a certain content, in this case a tensed content. An agent might represent an occurrence as both unpleasant and as in the past without thereby being in a state of relief. Rather, insofar as the state of relief involves a representation with a certain tensed content, what is crucial is the way in which the occurrence of a state with a certain tensed representational content is keyed to a specific phase of the motivational cycle. It is its occurrence in the motivational cycle, and the structure of the motivational cycle itself, rather than how the event in question is represented by the agent, that explains why that state of relief is one that can only occur once the occurrence in question is over.¹⁰ To this extent, this explanation of the perspective of relief is an instance of what in Chapter 1 I called the embedding strategy. The explanation makes essential reference to the actual temporal-causal structure of the sequence of events in which the agent's state of relief is embedded, rather than just to its representational content.

The second point is that the above style of explanation makes no particular mention of deliberation. It is no part of the explanation of relief's temporal orientation that, for instance, the agent has sources of evidence about the stimulus event which screen off any correlations between their present actions and its occurrence—although that

¹⁰Absent further elaboration, this account requires that, in the most basic case at least, relief is *factive*, in the sense that an agent's being relieved that an unpleasant occurrence is over implies that it is indeed over. The English construction 'X is relieved that _____' does appear to be *factive*; but nevertheless one may feel there is an intuitive sense in which the state of relief itself is not. For instance, I may feel relieved that my root canal operation is over (or so I think), when in fact the dentist is only just getting going. What I think the possibility of such cases indicate is that the state we describe as being relieved that a certain occurrence is over is in fact a much more complex state than the basic visceral response which typically follows the cessation of an unpleasant stimulus, but nevertheless retains an intrinsic connection to it. The operation of motivational cycles in sophisticated agents is mediated via structures of beliefs and longer-term goals, which introduce the possibility of relief being prompted by the receipt of information that turns out to be incorrect. For a nuanced discussion of the complexity of relief see Hoerl 2015a.

may well indeed be true. But there is a more primitive component to the temporal orientation of relief, one that stems from the fact that acting, and being motivated to act, is a psychological process that has a certain intrinsic temporal direction. Moreover it is a psychological process that involves conscious awareness: it is essential to the kind of motivational cycle I am considering that the agent, in acting and being motivated, is aware of what it is doing and what is going on around it—it occupies an agential point of view on its own actions and surrounding environment.

Matthew Soteriou, in a recent paper, suggests that having such a point of view, with its tensed orientation to the past and future, is characteristic of the kind of consciousness a creature typically enjoys when it is awake:

An account of that functioning synthesis [that enables awareness of a temporal span beyond the immediate present] is likely to depend crucially on the role played in it by the agential perspective that a subject has on what she is doing when engaged in action. That is to say, it is likely that an account of the temporal point of view associated with being awake will need to be an account of a temporal point of view that is both anchored by one's knowledge of what has just happened and shaped by the psychological posture to one's future that comes with an agential perspective on one's future—the sort of perspective on one's future one has in virtue of engaging in goal-directed intentional action, even if that action is as simple as that of exercising agency over the occupation of one's conscious attention. (2019, p. 131)

It should be clear that the kind of agential consciousness Soteriou has in mind does not necessarily require that the agent be exploiting dependencies between variables in its environment in order to pursue its goals. This is not at all obviously involved, for instance, in Soteriou's example of exercising agency over what one is attending to. The temporal orientation of such consciousness does however involve the fact that a conscious agent has a distinctive epistemic orientation to the past and to the future. A fuller account of the significance of this epistemic orientation would presumably involve the different ways in which information about the past and the future feed into the agent's action guidance. But such facts about the flow of information in agential consciousness again characterises a more primitive level of awareness than anything involving a grasp on the agent's part of causal relationships.

On the other hand, in characterising the perspective of a relieved agent in terms of

position of relief in the motivational cycle, I have made free use of causal notions. The key idea is that the motivational cycle constitutes a causal process in which the onset of the unpleasant episode prompts an aversive reaction and its later cessation prompts relief. Deliberation—understood as the selection of action routines guided by knowledge of exploitable causal dependencies in the environment—is just one aspect of the natural processes that make up the lives of animate agents. The temporal perspective of a relieved agent is due to the intrinsic temporal and causal structure of such processes.

This raises the question, though, how the temporal perspective associated with the agency of motivational cycles may be transformed in the case of agents who do consciously deliberate, as well as being subject to the more immediate impulses of and aversion and appetite.

In contrast to creatures who are modified only by simple animal desires, complex agents like humans are subject to a varied and multidimensional array of desiderative forces. As well as immediate appetites, sophisticated agents have general interests and values, like a love of music or a commitment to intersectional politics. Such interests give rise to preferences about how the world should be, at a remove from one's states and appetites and from what it is in one's power immediately to pursue. The range of things such agents can form preferences about is in principle indefinite, limited only by their cognitive powers of conceiving of the relevant subject-matter. One can have preferences about the distant past, or about higher mathematics, or about the outcome of the 2019 Ashes series. The typical expression of such preferences is, rather than the pursuit of their satisfaction, the motivation to *find out* whether or not they are satisfied; and the disposition to feel glad or disappointed on making observations which are congenial or disagreeable to them. This kind of affective response, unlike relief, is one that can be had in principle at any time: what occasions the response is the epistemic shift in one's beliefs or credences about the matter in question.

We should thus acknowledge two faces to desire: the motivational component, which moves agents to purposive activity aimed at attaining some desired object or state; and the evaluative component, which directs agents' epistemic activities, and dictates their affective responses to the news.

This is not necessarily to suggest that these two roles are necessarily played by dis-

ting desires. Many desires play both. For example, my preferences about the upcoming election might motivate me directly to attempt to influence its result, by voting and canvassing; but also to closely follow the political developments during the run-up, to consult pollsters, and so on, and to be glad or disappointed when I receive evidence of my favoured party doing well or badly. Such cases are common; their existence reflects the fact that we have interests and preferences that concern a wider world beyond our own states and actions, but which we are capable of intervening in and influencing to a limited extent.

The motivational and evaluative components of desire can interact in various ways. One acquires an interest in whether a certain state of affairs obtains if one believes that its obtaining is relevant to our getting what one wants. One such way in which agents acquire interests in states of the world very far removed from their agency is by making assertions, propounding scientific theories, and so on. These activities are effectively bets in which one is rewarded by having made a true statement if a certain state of affairs obtains, and penalised with falsity if not.

Conversely, almost all of the desires on which we act, even those simple appetites which we share with other animals, are informed, in one way or another, by our interests, and our sense of what is valuable and worth promoting. Most human interests beyond the immediate appetites are on the level of personal relationships, families, professional organisations, social and political institutions, and so on. These areas of concern map out a scene of action that we think of as beyond our power to influence unrestrictedly and without mediation; but at the same time as meaningfully susceptible to certain kinds of intervention, and as directing and organising our conduct.

The fact that we can be motivated by desires formed at different levels of abstraction from our immediate states and appetites also introduces a dependence of desire and motivation on belief. Which desires motivate us is determined not just by the urgency or vivacity with which they strike us, but with our assessment of the prospects of their satisfaction, beliefs about further consequences of pursuing some desires over others, and so on. The world that we act on and are affected by is the very same world that can find out about through our epistemic activities. As inhabitants of such a world, the challenge we face is to harmonise our actions and choices with the wider sphere of our

interests.

The conception of rational agents as maximising expected utility offers the prospect of a theory that systematically relates motivation, preference and belief. The formula for expected utility provides a measure of the desirability of an action in terms of its tendency, given the agent's credences, to promote those outcomes which the agent favours; and an agent who maximises expected utility is always motivated to perform an act which has at least as much expected utility as any other act feasible at that time.

The central desiderative notion on this conception is that of desire as preference among states of the world or propositions. Any proposition can be assigned a *news value*, recording how glad or disappointed an agent would be to learn its truth. The distribution of news values over different propositions is constrained by agents' beliefs about how they are evidentially related—by what provides news of what.

So far, this structure of rational belief and desire could in principle describe a subject incapable of acting, but able to observe and evaluate states of the world. For creatures that are agents, however, there is a distinguished class of propositions—those concerning the agent's own actions—which the agent can make true at will. That is, in acting, an agent can *give* themselves evidence of the truth of a proposition: they can 'make the news'. Thus the motivational role of desire is subsumed as a special case of its evaluative role.¹¹ The paradigm of this conception of belief, preference and motivation, and their relations, is the version of decision theory developed by Richard Jeffrey. Jeffrey is moreover explicit that the fundamental notion of preference is not intrinsically connected with agency: as he puts it, 'the notion of preference between propositions is neutral regarding the active-passive distinction.' (1983, p. 59).

The core of this conception of rationality is a structure of rational belief and evaluative preference. Consequently it is not an essential component of the theory that agents act and decide in time, and in being motivated are modified by occurrent desires. An illustration of this is that an agent in the grip of an occurrent desire will be represented as having a schedule of preferences which are *time-biased*. For instance, an agent under-

¹¹This of course raises the key question at issue between EDT and CDT, whether acts are to be evaluated in terms of their auspiciousness or their efficacy. The idea that the value of act-propositions is their news value might suggest a commitment to an evidentialist rule, insofar as news value is computed using Bayesian conditional probabilities. However Joyce (1999, 2000) convincingly argues that this is not the case, defining a notion of 'causalised news value' in terms of which acts are evaluated.

going a painful experience will prefer states of the world in which the experience is over by the present time to those in which phases of it still lie in the future. From a certain, temporally detached evaluative point of view, such a preference can seem arbitrary. It is hard to see how there could be anything intrinsically better about a possible world when a certain painful episode occurs at one time as compared with a world in which an episode of precisely the same length and qualitative character occurs at a slightly later time. Even given the various partial and contingent interests and concerns of the agent in question, it is hard to see what could motivate any such evaluative preference between worlds which differ only with respect to the temporal location of a particular painful episode. Nevertheless, as long as an agent is acting on an aversive desire, it will seem as though they do have such a preference just as long as we take motivation to be a special case of such evaluative preference. Moreover, the agent's preferences will seem to be constantly *changing* as time goes by. This can encourage a picture of a temporally extended agent as a collection of time-slice agents with preferences that differ with respect to the temporal locations of certain events.¹²

What is missing from this picture is the idea that motivation is an occurrent and dynamic phenomenon. There is nothing mysterious or arbitrary about the fact that agents prefer painful experiences to be in the past as long as one keeps in view the fact that such phenomena as aversion and pursuit are not, fundamentally, evaluative states, but rather active episodes in the life of an animate agent. This is not to say that the description of agents as time-biased is spurious. It is just true that, for an agent with an aversive desire, possible worlds in which the pain ends earlier will be preferable to ones in which it ends later. Moreover, such desires can, in certain intertemporal choice contexts, lead agents to make series of decisions which are self-defeating.¹³ The point is that a fuller picture of the way agents act in time must take into account that such

¹²This is very much the picture that emerges, for instance, in Part 2 of Parfit's *Reasons and Persons*.

¹³Specifically, agents with hyperbolic time-discounting functions—i.e. functions such that the ratio of discounting between two points in the remoter future is less than the ratio between two points the same distance apart in the near future—will tend to reverse their preferences over options, and can be exploited on that basis. For a comprehensive review of time-bias and time-discounting, see Frederick, Loewenstein and O'Donoghue 2002. It is harder to construct cases in which the simple preference that pleasant experiences be in the future and painful experiences be in the past—so-called 'future-bias'—leads to self-undermining choices. For an attempt to construct such a case, see Dougherty 2011; cf. also Greene and Sullivan 2015.

preferences are grounded fundamentally in dispositions to act, rather than to regard some courses of history as more satisfactory than others.

This point applies no less to other, less visceral, motivational states. For a desire to actually get one to act, one has to be disposed not just to regard certain news items with favour or disfavour, but also actively to pursue one's goals by moving one's body and interacting with one's environment in appropriate ways. Desire has to get one's limbs in motion, and this is, fundamentally, something that happens over time.

The key point I want to make here is that the active state of being moved by a desire is characteristically one in which one's thought and attention is modified in certain ways. This feature is obvious in the case of simple appetitive or aversive desires—part of what it is to be affected by such a desire is for one's conscious experience to be so modified. The generalisation of this feature, as I shall argue in the next section, is the way in which intention preserves and stabilises one's desires over time. This motivating role of intention is connected with the knowledge it involves of the desirability of what one is doing. This in turn will explain why, in some cases, there are thoughts about the causal explanation of one's action that become available only once one no longer has the relevant intention.

4.3 INTENTION AND KNOWLEDGE OF WHAT ONE IS DOING

Almost anything worth doing takes a certain amount of time and involves multiple phases. This is true no less of ordering a coffee than of writing a novel, or surveying the bat population of an area of woodland. Typically, one lacks the power to perform a simple act that will set the wheels in motion and see to it that some complicated endeavour gets pulled off without one having to do anything further oneself. Rather, getting something done requires continuing to be motivated to pursue it as one's end. At the same time, we are creatures of conflict, continually subject to desiderative forces that pull us in different directions. In order to do anything worthwhile, we need to be able to stabilise our desires in order to focus on completing a specific task. An agent with no means of doing this—a 'wanton', in Frankfurt's term—would find it hard to get anything

done.¹⁴

One way for agents to ensure the execution of their longer-term plans is to employ self-binding and pre-commitment strategies.¹⁵ This can consist in arranging the external circumstances of one's future choices to make certain courses of action impossible; or else manipulating one's future payoffs so as to make certain actions undesirable. For instance, someone attempting to give up smoking can pay someone in advance to forcibly remove any cigarettes from their vicinity, or can take a drug that causes them to feel nauseous immediately on tasting tobacco smoke. More normally, however, agents do not need to go in for these kinds of strategies, because they can be confident that the desires that motivated them to take a certain decision will continue to motivate them to execute it. The basis of this stability is the capacity to make decisions and form intentions.¹⁶

Thomas Pink expresses the motivation-perpetuating role of decision and intention like this:

Even if, prior to any decision, we are already motivated to perform a particular future action A, there is still the question of whether we should remain so motivated into the future. If how we shall act is to be settled in advance—as the coordination of our action over time demands—there must not be a possibility that our motivation will change. The point of taking a decision in these cases, then, is to determine action in advance by ensuring that we do remain so motivated. (1996, p. 124)

¹⁴I am not meaning necessarily to endorse Frankfurt's specific use of the term as someone lacking 'second-order volitions'. One does not have to endorse Frankfurt's picture, on which the regulatory role is played by second-order desiderative states, in order to recognise the kind of incapacity of being unable to regulate one's desires in accordance with longer-term goals.

¹⁵Varieties of such strategies are catalogued in Part 1 of Elster 2000.

¹⁶In what follows, I will be focused more or less exclusively on the motivational role of intention. I will be less concerned with the 'planning' role, emphasised by Bratman 1987, which involves, for instance, the fact that if one intends to do A one normally forms further intentions on the assumption that one will do A, and that one's intentions at a time are subject to means-end consistency requirements. These role of intention are at least notionally distinct, since one can imagine an agent whose underlying motivations are highly stable but with extremely limited planning capacities, and a capricious planning agent who constantly switches between elaborately worked-out plans. Although I find it plausible that the planning role and the motivational role are nevertheless importantly connected, I will not pursue these connections here. Accordingly, in what follows I draw no sharp distinction between having an intention for the future and acting with an intention, since as I see it there is no significant difference between the issue of remaining motivated from the point of taking a decision to the point of beginning to act, and that of remaining motivated throughout a period of intentional action.

Pink goes on to argue that decision's role cannot be to ensure that the relevant beliefs are maintained, since if a belief on which a decision is based is rationally undermined, that ought to undermine the rationality of the decision too. He thus concludes that decisions perpetuate motivation by stabilising desire:

The decision's function...must be to ensure the retention of the desires that motivated it. Decision-making functions as a *desire-stabilising* agency. For as long as the beliefs which have motivated a decision remain unchanged, the decision should ensure that the desires that motivated it remain unchanged also. And that way it can be true...that intentions possess *inertia*. Intentions tend to persist until the time of the actions intended. That means deciding to perform an action—the formation of an intention to act—ensures a persisting motivation to act as decided. Since motivation depends a priori on desire—since an agent can only be motivated to perform one of the those actions he desires most strongly to perform—stabilising desire helps stabilise motivation. Stabilising desire ensures that from the time of decision onwards, the agent retains desires that allow him to be motivated to act as decided. (pp. 124–125)

There is a weaker claim and a stronger claim in the vicinity here. The weaker claim is just that the function of decision is to ensure the motivating desire persists in some form (conditional on the persistence of relevant beliefs.) This is not enough to ensure the performance of the action, though, because we do not act on all of our desires. Indeed we most of the time have conflicting desires, so acting on them all is an impossibility. A stronger claim, then, is that the function of decision is to ensure that the motivating desire persists, and continues to motivate one. Only this stronger claim, it seems to me, secures that the function of decision is such as to ensure the subsequent execution of the action decided upon.

The more general idea here is that the act of deciding, and the state of intention that typically results from decision, is a psychological process which leads to action through its influence on what the agent is motivated to do, where this influence is reason-based in a way that contrasts with the more manipulative influence that can be exerted through the use of self-binding strategies.¹⁷

¹⁷As Soteriou 2013, ch. 12.4 observes, the influence that decision-making has on future conduct is importantly dependent on the agent's capacity to remember what they have decided. The kind of memory involved here is distinctive in that involves not just remembering that a certain mental episode in fact occurred, but doing so in a way that involves retaining a commitment to performing the act decided

In this respect the cycle of decision–intention–completion is very broadly analogous with the cycle of pain–aversion–relief discussed above. Like aversion, intention is an active state that controls an agent’s behaviour by influencing their motivation. In the case of aversion, the influence on an agent’s conduct manifests in certain immediate modifications of the agent’s conscious thought and attention. For instance, the painful stimulus is highly salient, and constantly directs the agent’s attention towards the need to get away. The presence of an intention need not modify an agent’s conscious state in the same manner, of course. Nevertheless we can see intention more broadly as modifying the agent’s motivational system, in ways that have less to do with conscious experience and more to do with the kind of self-knowledge one has in intentionally acting. But, as I will argue, the underside of this self-knowledge is a kind of self-ignorance about the psychological causes of one’s action.

As Anscombe famously pointed out, whenever one is doing something intentionally one is able to answer the question *why* one is doing it, and do so without pausing to introspect. The point of Anscombe’s observation is that acting intentionally means having *practical* knowledge of change happening in the world, knowledge which is distinctive in being active rather than passive with respect to what it is knowledge of. On the face of it, the idea that one has a distinctively practical knowledge of one’s own actions would seem to be in tension with the idea that there can be a causal insight that becomes available only after acting. However, as I shall argue, these two kinds of knowledge—practical knowledge of what one is doing, and retrospective knowledge of the causes of one’s action—are importantly different.

The practical knowledge of one’s own actions illuminated by Anscombe is intimately connected with the structure of means-end practical reasoning. An answer to Anscombe’s question ‘Why are you X-ing?’ typically mentions some end that X-ing will enable the agent to achieve. The answer can take a variety of different grammatical forms, and it may be elliptical, requiring some background knowledge about the agent and their world in order to make sense; in general, though, the answer works by

upon. This distinctive kind of remembering may be signalled in English by the infinitival construction: *remembering to do something*. (Compare with the infrequently used, but perfectly legitimate, construction, *knowing to*.) This dependence on memory seems to be key to understanding the sense in which the influence that decision exerts over future conduct is non-manipulative.

specifying the point, or the good, that the agent sees in X-ing.

Whether an answer to Anscombe's 'Why?'-question amounts to properly causal explanation has been much discussed, and I do not wish to add to that debate. Certainly a proper answer to the 'Why?'-question is importantly *related* to causal explanations. It would be informative to someone writing the agent's biography, for instance. What I want to stress, though, is that an answer to the 'Why?'-question leaves crucial bits of causal information undetermined. In particular, it does not explain why the agent chose X-ing as a means to their end, rather than some other available means. It also does not explain why they chose to X at that particular time; or why, at that particular time, they chose to pursue the end furthered by X-ing, rather than some other end of theirs. Sometimes this may be obvious: it may be that at that time, X-ing was the only salient option that would further any of one's ends. In other situations, though, this is exactly the kind of thing one can become retrospectively puzzled by. One can say to oneself, 'yes, but why exactly did I choose to do X then, when previously I wouldn't have thought of X-ing at all?' The fact that there was a certain amount to be said in favour of X-ing is only the beginning to answer of this kind of question.

Moreover this is not just information that is unspecified but could in principle be added; it is a structural feature of the kind of information provided in an answer to the 'Why?'-question that it is left unspecified. The kind of information provided by an answer to the 'Why?'-question is essentially *non-contrastive*. It just says what is good or desirable about an action; it says nothing about why that particular action should be performed over other desirable actions.

Anscombe makes the point about the non-contrastivity of practical knowledge in connection with her discussion of the practical syllogism. The general form of the practical syllogism is:

I want X/X is desirable.

If I do A, then X will obtain.

So, I'll do A.

The distinctive and puzzling feature of such transitions is that they cannot be construed

as expressing anything like a logical consequence relation. There is no premise of the syllogism which asserts that X will not obtain if I do not do A. The kind of relation of the premises to the conclusion is rather one of *sufficiency*—the premises state practical grounds which are enough to motivate the conclusion, without necessitating it. This property of the practical syllogism as expressing a sufficiency relation is illuminating of the character of an agent's knowledge of why they are doing what they are doing. In intentionally acting, one understands one's activity in terms of the features that make it worthwhile. But this need not imply any kind of necessity to one's acting on that motive in that way, nor even any comparison with alternative, equally well-motivated courses of action.

The kind of action explanation provided by the practical syllogism is thus extremely modest compared with that provided by the conception of rational agents as expected utility maximisers. The rule of utility maximisation deals with a generalised conception of value with respect to which diverse ends can be compared, and presents a formal theory of how to calculate the degree to which different courses of action promote it. Selecting a course of action that fails to maximise expected utility is always rationally impermissible. Thus an agent who maximises expected utility, and knows themselves to do so, knows considerably more about the causes of their action than an agent who just possesses an answer to the 'Why?'-question. Such an agent knows precisely enough what their utilities and credences are and to be confident that the course of action they are pursuing is optimal given those utilities and credences.

The idea that an agent might decide on a course of action by *first* consulting their utilities and then engaging in a processes of reasoning that approximates to an expected utility calculation, though, is highly unrealistic. What a utility function measures is very different from any intuitive notion of desire. Utilities are defined over *outcomes*: total states of the world in which variables for everything that the agent cares about are given a specific value. For real agents with multidimensional interests and concerns, the utility of an outcome will be a complex alloy of its diverse good- and bad-making features. Evaluating the utility of different possible outcomes, then, requires being able to survey the various subjective sources of value and apply them, in some principled way, to a range of possible situations in order to compare their desirability. Surveying

one's values in this way would mean adopting a kind of disengaged stance towards one's own subjective attitudes, in order to assess a possible situation with respect to different, possibly conflicting, dimensions of evaluation that one's attitudes determine. Perhaps this is sometimes possible. It may be that in certain cool hours one can reflect in such a disengaged fashion on how satisfied a certain total situation would make one overall. But, as I argue in the next few paragraphs, this disengaged stance is incompatible with the way that desire motivates when one acts intentionally.¹⁸

The non-contrastive character of the practical knowledge one has of one's reason for acting is connected with intention's motivational role. In acting intentionally, one's attention is occupied by the end that makes the intention worth pursuing. Intention preserves motivation because, in having the intention, one is not disposed to reflect on the other ends one might pursue, or on the other ways one might pursue that same end. To the extent that one is so disposed, one's reflection on the other alternative courses of action available undermines the extent to which one is actually motivated to do as one intends.

This is most obvious in cases where doing as one intends involves some kind of concentrated activity, like writing an academic paper. In such cases, to the extent that one is disposed to reflect on the desirability of alternative ways of occupying one's time, one ceases to be motivated to continue with the activity of writing a paper; if one is

¹⁸Note that, if what I am arguing is correct, this does not necessarily imply there is anything wrong with the idea that a fully rational agent maximises expected utility. It only undermines the idea that one might realistically understand oneself as maximising expected utility on the basis of knowledge of one's ultimate utilities. I am thus largely in agreement with Joyce when he writes, 'No sensible person should ever propose expected utility maximization as a decision procedure, nor should he suggest that rational agents must have the maximization of utility as their goal...The expected utility hypothesis is a theory of "right-making characteristics" rather than a guide to rational deliberation. It in no way requires an agent consciously to assign probabilities to states of the world or utilities to outcomes, or to actually calculate anything. The decision maker does not need to have a concept of utility at all, and she certainly does not have to see herself as an expected utility maximizer. The demand is merely that her desires and beliefs, however arrived at, should be compatible with the expected utility hypothesis in the sense that it should be possible for a third party who knows her preference ranking to represent it in the way described.' (1999, p. 80) To this we can perhaps add that utility maximisation may be a useful goal in specific, limited cases where one's utilities are well-defined and known—for instance, decisions where only monetary payoffs are involved—and one may thereby know one's decision to be utility-maximising. Likely, it may be that one can apply the notion of utility maximisation in order to detect local inconsistencies in one's preferences and beliefs. The point is just that it is highly unrealistic to suppose that, in general, one might know one's decision to be rational on the basis of having consulted one's utilities and credences, and performed a subsequent calculation.

systematically and repeatedly distracted in this way, it is no longer the case that one is actually writing a paper. But the same is true of intentions whose possession and execution does not require concentrated activity in the same way, like the intention, had on Monday, to go to the cinema on Wednesday. Insofar as one is actually motivated to go to the cinema on Wednesday in the manner characteristic of intention, one is not disposed actively to consider the attractions of alternative incompatible courses of action.

The point is that having one's motivation preserved in the manner characteristic of intention involves a kind of selective attention to the desirability of the courses of action available to one. But, in general, the way one finds out what one wants, overall, is precisely by reflecting on one's options and considering how attractive they seem. The result is that the kind of commitment or motivation-perpetuation involved in intention imposes a kind of limitation on one's ways of coming to know about the subjective sources of one's action. Consequently, the contrastive causal why one chose one particular course of action over another, both of which one had some reason to do, is, in many cases, a question that cannot be raised without disturbing or interfering with the sources of motivation it concerns. And this is precisely the kind of thing that can become clear in hindsight, when one reflects and tries to make sense of one's actions.

4.4 THE VALUE OF REGRET

We are not just saddled with a schedule of preferences over states of the world, which we then have to figure out how best to promote. Rather, the fundamental source of most of our preferences is our nature as agents, which manifests in dynamic motivational processes. In order to be happy, we have to be able to rise above the pushes and pulls of our immediate urges and desires, and moderate our conduct in accordance with longer-term goals and values. But we can only do so to a limited extent: being subject to the grip of such motivational processes is the essence of our condition as changeable agents. There is no stable resting-place from which we can stand back and coolly survey all our desires at once, because being motivated by a desire is precisely to be disposed not to treat it in that detached way. Being motivated to act thus involves a kind of blindness to

the conditions of conflict under which one so acts.

It is in the context of this predicament that we can see the distinctive value of retrospective attitudes like agent-regret. In the first place, regret can usefully be brought to bear on decision-making. One commonplace way in which people decide whether to take some course of action is by asking, not, ‘Is this the best thing to do, all things considered?’, but rather, ‘Will I regret (not) doing this?’ One might distinguish two ways in which regret can play a role in decision making, as ‘experienced’ and ‘anticipated’ regret.¹⁹ In the first case, the idea is just that, if one makes a certain choice and regrets it, one will then avoid making the same choice in the future. This role of regret in decision-making is relatively uncomplicated, and there is evidence that young children can reason in this way.²⁰ All one need assume about the agent to make sense of this motivation is that experience of regret functions as a kind of signal that one made a bad choice, and if it becomes associated with that choice can thereby serve as a warning against similar choices in similar situations.

Anticipated regret, by contrast, involves imagining the retrospective perspective one might alter occupy on one’s decision. This is much more complicated: it requires thinking oneself imaginatively into the position on one’s choice that one will occupy in the future, if one takes it, and taking one’s imagined affective response as a guide to one’s present decision.

The fact that anticipated regret can be motivational and action-guiding, though, might seem to involve a kind of puzzle. Someone might argue: either one’s later self is anticipated to have the same preferences as one has now, or different preferences. If one’s later self has the same preferences, then the decision in question should be an equally bad idea given one’s preferences now. It is hard to see why one would go to the trouble of imagining a later perspective in order to work out that the action is not desirable. On the other hand, if one’s preferences will be different, it is unclear why one should be motivated by them. If some end seems worth pursuing now, then why should one be moved by the fact that it will not seem to have been worth pursuing later? This question can be raised in the case of the time-biased preferences generated by simple

¹⁹These terms are from Hoerl and McCormack 2016.

²⁰See for instance O’Connor, McCormack, Beck et al. 2015; O’Connor, McCormack and Feeney 2014 for studies of how regret affects young children’s decision-making.

appetitive desires, like when I eat all the crisps now, knowing I will be hungry later; or cases involved anticipated wholesale value change, like Parfit's (1984, p. 327) Russian nobleman, who believes that his socialist ideals will have faded by the time he inherits his fortune, and so enlists his wife to hold him to commitment to giving it away. In such cases, one is confident that one will regret one's decision. But, because one does not share the preferences of one's future self, this thought may have little motivational force.

The present discussion provides a way to understand how anticipating the future perspective of regret can be illuminating with respect to the choice one currently faces, and so can be choice-guiding. It may be that, because of the perspective one occupies on a decision as agent, one is unable fully to step back and obtain a clear view of the desires that one is in the grip of. Thus reflection alone is unable to guide one to a fully considered decision. In such cases, one may project oneself into a subsequent future perspective that one anticipates occupying on the decision, and thereby gain an understanding of one's present situation that otherwise would have been unavailable. It may be clear from a later perspective that one would be acting only out of anger; or that one would be foolishly passing up an opportunity to do something one has always wanted to do. In having these thoughts, one is not arbitrarily attaching greater importance to the satisfaction of a later time-slice's preferences than to one's present desires; rather, one is attributing a kind of authority to one's later self with respect to what it is one really wants, and to how one's current motives fit into a wider system of concerns and interests.

This operation is quite psychologically complex. It involves, both a capacity for imagination, and the ability to conceive of the imagined later self—the subject of the imagining—as possessing a special insight into one's present situation that one now lacks. Noting this psychological complexity leads to a point about the broader function of regret, beyond the pragmatic value of anticipated regret for decision-making. Regret—along with positively valenced retrospective attitudes, like pride—involves the recognition of past occurrences as distinctively one's own, and so serves to support a conception of oneself as a temporally extended agent. This kind of extended self-conception is summarised by Charles Taylor, in a characteristically suggestive passage:

I don't have a sense of where/what I am...without some understanding of how I have got there or become so. My sense of myself is as a being who is growing and becoming...as a being who grows and becomes I can only know myself through the history of my maturations and regressions, overcomings and defeats. My self-understanding necessarily has temporal depth and incorporates narrative. (1989, p. 50)

The point I want to insist on is that the kind of temporally extended self-understanding which Taylor is articulating here has an irreducibly causal dimension. The understanding of oneself as persisting over time, whatever else it involves, includes the idea that one's life as an agent is, among other things, the causal history of particular living individual: one's actions have causal roots in one's desires and passions, and the 'maturations and regressions' Taylor refers to are grounded in the causal processes of motivation and action. Picking up on a phrase used in earlier chapters, we might think of this aspect of self-understanding as involving a grasp of one's life as an agential route through the world. Moreover, this causal self-understanding is not fundamentally articulated in terms of the causal information that might figure in the stance of deliberation. Rather, what it involves is essentially a special case of what in the previous chapter I called a historical use of causal concepts, as applied first-personally to one's own agency. This is, fundamentally, a certain kind of *explanatory* perspective on the world, one where certain changes are identified as distinctively one's own work, and evaluated in that light. The next chapter will consider some of the cognitive underpinnings of this explanatory perspective.

This last point, however, raises a puzzle that I think is already latent in Williams's idea of agent-regret. In saying that the retrospective perspective is an explanatory perspective, in a way that contrasts with a deliberative perspective, can we continue to maintain that it is an agent's perspective? The perspective of agency might be thought to be essentially prospective, distinguished precisely in being the kind of attitude one has to some event when one is deliberating over it. If we think of an agent's perspective in this way, then past exercises of one's agency are just more events in the history of the world—albeit events to which one often has some special and first-personal form of epistemic access in autobiographical memory. Something like this background thought can, perhaps, be detected in Wallace's resistance to the idea that there is a specially agential way

of relating to one's past occurrence that is not grounded in a sensitivity to considerations that one could and should have been sensitive to at the time of acting.

I think we can see this same puzzle underlying the tension between deliberative and historical uses of causal concepts explored in the previous chapter. The deliberative paradigm presupposes a conception of oneself as, if not outside the world like Wittgenstein's metaphysical self, then at least causally exogenous to it. This attitude is captured in F. P. Ramsey's famous remark that 'my present action is an ultimate and the only ultimate contingency.' (1978, p. 146). More recently, according to Jenann Ismael, 'from the self-centred perspective of a system whose activity is *part* of the pattern [as opposed to a God's-eye view 'outside' the pattern of events], its own actions necessarily have the status of interventions.' (Ismael 2007, p. 5) But this conception must break down, it seems, once one self-consciously recognises oneself as an element of the causal order, as one necessarily must when reflecting on the causality of one's past actions.

One can get into a frame of mind where this agential self-recognition seems unintelligible. For if the first-person perspective is indeed a perspective from which one's actions are necessarily interventions, then how could the identification of any item in the causal order—an animal, perhaps—be an identification of an agent as *oneself*

Against this, it is important to remind ourselves that the thought of oneself as a participant in the causal order is not in fact alien to ordinary thinking, and that this comes out most clearly in thought about oneself in the past. Despite theoretical reservations about them, attitudes like agent-regret, and other similar forms of retrospective assessment, are an ordinary part of our affective and ethical lives. This fact should enjoin us to consider the idea that the kind of perspective one has on one's past actions is nevertheless an agent's perspective; accounts of the perspective of agency, self and causality that take the deliberative paradigm as primary are liable to miss this out. Perhaps there is a deep puzzle about how this can be so—the puzzlement expressed in Nagel's question how anything in the objective order could be identified as me. Insofar as this puzzle is in play, I do not claim to have the materials to answer it. The point I am pressing is just that ordinary self-consciousness does include a recognition of the causal structure of one's past agency.

5 | TEMPORAL COGNITION AND THE CONCEPT OF THE PAST

In the two previous chapters I argued, in different contexts, for a distinctively retrospective perspective that agents have on their actions. I claimed that this retrospective perspective involves a kind of causal thinking that is not fundamentally concerned with agentially exploitable relationships. Rather, the retrospective perspective involves a distinctive kind of knowledge one has of one's agential route through the world. In this chapter I want to look in more detail at the ways of thinking about time and causality, and their interrelations, that are involved in this retrospective perspective.

This chapter will put the historical use of causal concepts in the context of the background framework of temporal thinking that it goes with. The key claims are that 1) the singular causal claims involved in historical explanation are distinctive in relating to temporal particulars—occurrences whose particularity is connected with their having a specific, single temporal location; 2) such causal claims also involve an implied contrast between the actual and the counterfactual, a contrast which is not implicit in generic causal claims; and 3) these two features are both related to a background conception of time as comprising a system of particulars. I will develop these claims in section 5.1 in comparison with alternative, less sophisticated ways of representing time, discussed both by John Campbell and by Christoph Hoerl and Teresa McCormack. In 5.2 I will consider how the background conception of time is connected with, and supported by, an adequate understanding of causal structure. In the final section I will make some speculative remarks about how this causal and temporal understanding is connected with a conception of one's agency as extended in time. This will connect with themes

that emerged from the previous chapter's discussion of agent-regret.

5.1 TEMPORAL FRAMEWORKS: SCRIPT TIME, PHASE TIME, AND GLOBAL TIME

Saying that historical causal explanation relates to temporally particular occurrences might seem like just a way of saying that what is in question are token- rather than type-level causal claims. In a sense this is correct. It is true that, the way we in fact draw the type/token distinction, token causal claims explain particular historical occurrences. What this obscures, though, is that drawing it in this way depends on a background temporal framework. We can bring what is distinctive about this by seeing some ways in which a thinker's understanding of time could be lacking.

In work carried out in the 1980s and 90s, Katherine Nelson and others (Fivush, Hudson and Nelson 1984; Nelson 1986, 1988, 1996), presented considerable evidence that a large part of young children's knowledge of the world is organised around *scripts*. The basic idea of a script, derived from artificial intelligence work by Schank and Abelson (1977), is that of a generic, repeatable pattern or sequence of events. Nelson describes a script as 'an ordered sequence of actions appropriate to a particular spatial-temporal context and organized around a goal.' (1986, p. 13) Scripts have certain key characteristics which make them particularly apt vehicles for children's learning: they are functionally organised in terms of specified roles and goals; they are hierarchically organised, with the potential to include subscripts or be subsumed within larger scripts; their structure is more or less flexible, and may permit decision points, optional subevents, permutations of order, and so on; they have the potential to incorporate causal connections between elements of the script; and their basic components are human actions. Some typical examples of children's scripts in modern Western societies might be the school day, the morning routine, trips to restaurants, birthday parties, and so on.

The main striking finding of Nelson and colleagues was that children of preschool age (i.e. 3–5 years old), who are known to perform poorly on more abstract Piaget-style temporal reasoning tasks, have surprisingly rich and detailed scriptlike knowledge of what happens in specific, socially defined settings, and can produce coherent verbal re-

ports of this knowledge when prompted by adults. There is thus good reason to think that script knowledge constitutes an important organisational framework for young children's thinking about time.

Clearly, there is a sense in which script knowledge is temporal, since it concerns events as they unfold in time. But knowledge that is entirely script-based falls short of the kind of temporal thinking that mature adult humans engage in. Specifically, script-like knowledge does not seem sufficient for thought about particular times or events remote from the present.¹ Adult temporal thinking, by contrast, can reach out to particular events the past in episodic memory; and we can refer to remote time and events descriptively, with the aid of calendar and clock systems. It is worth exploring just the sense in which scriptlike thinking fails to relate agents to particular times.

Script knowledge goes together with a kind of tensed thinking. A script is not just an abstract representation, but something that an agent puts to use in making sense of the world around them. In order to be able to make use of a script, one needs to be able to recognise instances of the script in one's immediate experience, and to keep track of its progress. Thus Hoerl and McCormack write:

...in order for children to make use of representations of repeated event sequences, they need to be able to know where they are within an unfolding sequence. That is, they need to be able to use their representation to orient themselves in some way so that they have expectations about what will happen next (e.g., putting on of pyjamas happens next), and can assume that events that have already happened are now over (e.g., once pyjamas are on, it would be very surprising to be given a

¹It is an open question whether script knowledge exhausts young children's understanding of time, or whether they also have some knowledge of particular past events. Various studies show children as young as 2 years being able to refer to unique and novel past events, e.g. Eisenberg 1985; Fivush, Hudson and Nelson 1984; Hamond and Fivush 1991; Reese 2002. However recall is typically patchy and heavily dependent on adult prompting and scaffolding. See Fivush 2011, pp. 565–568 for review and discussion of some of these data. Moreover, although children will report events using the past tense, some psycholinguistic studies (e.g. Antinucci and Miller 1976; Wagner 2001, 2012) suggest that children's mastery of the tense system is only partial at this stage and does not evidence an ability to think of particular past times. Clearly there are obstacles here to assessing the extent to which young children's reports of past events are partial due to a lack of an ability to think about the past, or more a lack of linguistic means to express their knowledge. There are thus complex empirical and theoretical issues surrounding the attribution to young children of the ability to think about particular past times in addition to possessing scriptlike knowledge. Nevertheless, it is relatively clear that, in early childhood, script knowledge plays a dominant role in organising children's thinking about time, in contrast to a more incipient ability to think and talk about unique past events. This suggests that script-based thinking and thinking about particular times are not just conceptually distinguishable, but actually developmentally separate abilities.

bath). (2017, p. 308)²

This suggests that we should think of children's script knowledge as naturally going together with a kind of tensed thinking, in which phases of a script are located in relation to the present moment by being labelled as past or future. Labelling phases in this way has concrete consequences for what the user of the script can expect to happen, or has the opportunity to do. For example, a child might locate themselves within their script of the morning routine by thinking of the getting dressed phase as past and breakfast as future, and so expect sitting down, opportunities for eating cereal, drinking juice, and so on.

This kind of location within a script provides an important precursor to the ordinary idea of the past as fixed. Script-phases located in the future are characteristically anticipated, whereas phases in the past are no longer to be expected. A further development of this basic idea comes if we imagine a script user as equipped with a basic grasp of how elements of the script might depend on its actions. If an agent thinks of its actions and the phases of the script as being related in systematic ways, we can see how it might thereby associate *future* with *possible to affect*, and *past* with *no longer possible to affect*.

An agent who thought in this way would thereby attach an immediate practical meaning to the notions of past, present and future. Indeed Hoerl and McCormack suggest that this understanding of the tenses is fundamentally *aspectual*, in that events are categorised in terms of their status as either completed, ongoing, or anticipated (1999, p. 166, p. 308). This use of the tenses is importantly different from their use by mature adults, in that there is no overarching timeline in which events are located.³ We can think of the script as a reusable *schema* that the agent uses to make sense of the

²Hoerl and McCormack make essentially the same point in an earlier discussion; cf. 1999, p. 162.

³Some evidence for young children's thinking about the past in this way comes from developmental linguistics. According to studies by Antinucci and Miller 1976, children's early use of the past tense is limited to achievement and accomplishment verbs (in Vendler 1957's terminology), with use of the past tense forms of stative and activity verbs occurring later. The hypothesised explanation for this is that early understanding of the past tense is limited to an understanding of past events in terms of their results in the present; thus only verbs which imply a result are used in the past tense. However some of the empirical evidence is disputed e.g. by Weist 1986. Weist claims that children's early use of tense does show an ability to locate tense in the past, but does not display Reichenbach 1980's distinction between 'event time' and 'reference time', as children do not use the past imperfective form of accomplishment verbs. On this hypothesis children's use of tense, despite locating events in the past, does not show an

world around them by categorising events aspectually into past, present and future; but different uses of scripts may fail to be integrated into a common framework.

Thus, someone who locates phases of a script as past or future need not think of the past phases of a script as fixed in anything like the way that adults think of the past as fixed. There is nothing in the thought of a script phase as past which rules out the possibility that the same phase might occur at a later time and be subject to intervention. For example, in thinking of the ordering phase of a restaurant trip as in the past, nothing rules out the possibility that I might order differently another time.

It is important to be clear about this point. The point is *not* that a script-using creature understands particular past events to be changeable. On the contrary, I am assuming that, as long as such a creature represents an event as past, it will thereby represent it as no longer possible to affect. But this representation is local to the use of the script in a given context, and so persists only as long as that particular use of the script continues. The fact that there may be another, subsequent instance of the same script belongs to an external, theorist's description of the agent's situation; the agent themselves, we are supposing, fails to distinguish between distinct instances of the same script, and so has no room for the idea that the now-past event might later be future. This illustrates the crucial point, made above, that representing the past as fixed is not just a matter of failing to represent it is changeable.

The point can be brought out in a different way by noting that thinking of the past as fixed involves an implicit contrast between what was and what might have been. If one believes that a certain event occurred, then the thought that it might not have occurred is, necessarily, counterfactual: it involves consideration of a scenario that is understood specifically to diverge from the actual course of events. By contrast, possession of a script provides for a kind of modal thought, but one that falls short of the properly counterfactual thought that goes with mature thinking about the past. A script may

understanding of past times as providing an alternative temporal perspective to the present. Note that both of these hypotheses about early linguistic understanding have only an indirect bearing on the hypothesis mentioned in the main text, that script-based tensed thinking is aspectual in the sense that parts of a script are classified quasi-aspectually as completed or to come in a way that gives them an immediate practical meaning. The practical meaning need not be encoded lexically (as with the verb 'break', but not with the verb 'run') but rather at the level of the child's knowledge of what comes next in the script. For further discussion see (1999, pp. 162–166).

contain a number of variables, decision points, and so on. In this sense, mastery of a script may involve recognising that there are a number of possible ways in which the script may be realised. Moreover, an agent that locates itself in a script may recognise that things could go differently from how they are actually going on that occasion. That is, they may have both tensed thought 'It's F, and X is happening' consistently with the generic thought 'When it's F, Y can happen'. But this alone does not amount to properly counterfactual thought. The idea of a possible realisation of the script is neutral between a way things might go on a different occasion, and a way things could have gone on this very occasion. Drawing this distinction requires recognising that there may be other, non-present, but equally actual, instances of the script in which one is presently located. And this recognition is not required for basic competence with the script.

The fundamental point is that someone might use a script without ever having the cognitive means or basis for distinguishing distinct actual instances of the same script. The fact that they do not so distinguish them means they cannot think of the present script-instance as located in a wider temporal structure in which other script-instances are located. And this in turn means they lack the resources for drawing a distinction, among possible configurations of the script that are not presently instantiated, between the actual and the non-actual.

I want to contrast my way of drawing the difference between a script-using agent and one that has a full grip on the concept of the past with some slightly different accounts, given by John Campbell and by Christoph Hoerl and Teresa McCormack.

John Campbell in *Past, Space and Self* makes a distinction between 'temporal orientation with respect to phase' and 'temporal orientation with respect to particular time'. He writes:

Consider an animal that hibernates. Through the part of the year for which it is awake, it regulates its activity depending on the season. Such an animal certainly has a use for temporal orientation. It can recognise that it is now late spring, perhaps by keeping track of how long it has been since winter, and realize that it will soon be summer. But it may not have the conception of the seasons as particular times; it may be incapable of differentiating between the autumn of one year and the autumn of another. It simply has no use for the conception of a particular autumn, as opposed to the general idea of the season. So while this animal is capable

of orientation with respect to phases, it is not capable of orientation with respect to particular times. (1994, p. 38)

Phase orientation differs from script knowledge in that the phases are regularly repeated, and typically are natural cycles rather than socially defined activities. But the key element they have in common is that a creature who thinks just in this way need only locate itself with respect to recurrent features, rather than particular temporal locations. And, while many non-human animals do indeed display complex timing behaviour that implies an ability to orient themselves with respect to natural environmental cycles (cf. Gallistel 1993), there is little evidence of animals identifying specific past times.⁴

Elsewhere Campbell notes that a creature who just uses phase time will not think of the past as permanently fixed. However, Campbell's statement of this point is liable to invite a subtle misunderstanding of the issue.⁵ This is how he puts it:

When you and I speak of events in the past...there is a sense in which their temporal location alone renders them insusceptible to being affected by us...No such conception of the past is available to an agent with only the conception of time as phase. Suppose...that I am able to affect what is for breakfast. Whether there is marmalade for breakfast at 10.00am is then something that I can affect. Just after breakfast on one morning, I can think, 'breakfast is just over'. Early the next morning I will be able to think, 'Breakfast is just about to come up'. But the only conception I have of there being marmalade for breakfast is the conception of a state of affairs I am currently able to change. I can think of there being marmalade for breakfast at 10.00am; but that is the state of affairs I am currently able to change. I can affect whether there is, in general, marmalade for breakfast at 10.00am. I can't now affect whether there was marmalade for breakfast yesterday at 10.00am. But that conception of a particular, unrepeatable time, yesterday at 10.00am, is just what is not available to an agent who has only the conception of time as phase. (2006, pp. 3–4)

⁴In an experiment carried out by Clayton and Dickinson 1998, scrub jays were shown to be sensitive to the amount of time elapsed between an event of some food being cached and the their return to the caching site. Clayton et al offer this as evidence that of the scrub jays retaining episodic memory-like information about the particular caching event. However, Suddendorf and Busby 2003 object that the scrub jay's behaviour does not necessary imply that they were locating themselves with respect to a particular past event, merely that the way in which they retained information about the cached food was sensitive to duration. The principal difficulty here is that it is very hard to imagine a clear indication of the ability to think about particular times other than the use of tensed language. The study and surrounding controversy is discussed extensively by Hoerl and McCormack, e.g. McCormack 2001, 2014; McCormack and Hoerl Forthcoming.

⁵I am not attributing this misunderstanding to Campbell, and I suspect he would disavow it. But the misunderstanding seems to me to be at least consistent with his way of putting things.

This remarks seem to imply that an agent who thinks in phase time will *always* take it to be within their power to affect some repeatable variable, like what there is for breakfast, so long as they are thinking about that variable at all. By contrast, I want to stress that, after some event on a given day, an agent that successfully orients itself with respect to that phase will, at that point in time, no longer think of that event as something it can affect. That point in the cycle is not one that affords opportunities to affect what is for breakfast. So, contrary to what Campbell's statement of the point implies, a phase-based agent does have *some* sense that the temporal location of an event renders it insusceptible to being affected. This is how such an agent is able to attach immediate practical meaning to the labels past, present and future.

Nevertheless, there is a sense in which Campbell is right that the only conception the agent has of there being marmalade for breakfast is of something which can be changed. This is the sense in which the agent's generic grasp of the cycle includes a grasp of various *possibilities* for breakfast; and their grasp of the possibility of marmalade for breakfast, on a day on which there was no marmalade for breakfast, does not have the character of a counterfactual. It does not amount to a grasp of how things could have gone, but did not go, on that particular occasion. But nor does it amount to a grasp of the possibility of its actually being able to bring about their being marmalade for breakfast tomorrow, if breakfast is not yet understood as being in the offing. Since the agents fails to distinguish occurrences of the cycle, it may have no grasp of *tomorrow*. The point is that the generic possibility of marmalade is not differentiated into a counterfactual thought and a thought about tomorrow. This means that a script-using agent might have, both, a grasp of the generic possibilities encoded in the script;*and* a grasp of the present affordances of their environment, one which might be expressed through an aspectual use of the tenses, but nevertheless fall short of properly counterfactual thought about the past.

Campbell's failure to emphasise these points is related to the way he conceives the underlying representation of phase time. He describes an agent who thinks in terms of phase time as having temporal notions like past and future, earlier and later, defined over 'a cyclically ordered domain of times' (p. 2). If this were the case, then such an agent would think of every point in the cycle as both earlier and later than itself, as both past and future. But this does not seem like the correct way to describe the temporal thought

of an animal timing its behaviour to fit with environmental cycles. A hibernating animal might locate its hibernation in the past, insofar as this means making use of information about the time elapsed since it woke up in calculating the time of year. But there is no particular reason to think of the animal as *simultaneously* locating its hibernation in the future. Perhaps we should do so if it started preparing for next year's hibernation as soon as it woke up from this year's—gathering materials for a nest, storing up food, and so on. But it would surely be more appropriate to describe this as locating *waking up* in the past and *going into hibernation* in the future—events with quite different causal meanings and affordances for the animal. Even in a case where we have grounds for attributing an animal thought of an event of exactly the same kind in both its past and its future, it is unclear what we would mean by saying that the animal *identifies* these events as one and the same. Crucially, the identification does not follow from the fact that the animal fails to distinguish these events. Rather, it may just not give any thought at all to the identities of particular events.

We can make a similar point about script time. An agent who employs scriptlike representations of events does not thereby operate with a nonstandard topology of time, for instance a branching topology. Rather, it is simply that the agent fails to recognise any temporal relations between different script-instances. Their conception of time is just incomplete, rather than nonstandard. A partial representation of order is not the same as a representation of a partial order.

The problem with characterising phase time as a cyclical ordering is that it credits the thinker's conception of time with more determinacy and systematicity than it should. Cyclical time is, as it were, a cosmic hypothesis about the global structure of time. Moreover, it is not immediately obvious that adopting a cyclical conception of time would prevent one from entertaining counterfactual thoughts about what goes on at particular points on the circle. By contrast, the point about an agent that thinks just with scripts and phases is precisely that they lack any overarching conception of time, of how different parts of the world they experience fit together into a system. Again we might suggest that, insofar as a creature using phase time has role for tensed thinking, their grasp of the meaning of the tenses is primarily aspectual: the significance of an event's status as past, present, or future lies in the expectations and affordances as-

sociated with its having that status. If this information is just discarded after a while, for instance when the phase begins again, there are no grounds to attribute to such a creature a conception of a common timeline uniting its disparate experiences, whether cyclical or linear.

Hoerl and McCormack, in their (Forthcoming), go to the opposite extreme, claiming that there is no compelling reason to credit non-human animals with the ability to think about time at all. They claim that the kinds of timing behaviour that Campbell explains in terms of phase time can instead be explained by positing a ‘temporal updating system’. In contrast to the ‘temporal reasoning system, which they hypothesise is only found in mature humans, the updating system does not explicitly represent any temporal information. It just represents information about the agent’s present environment, updating these representations as time goes on. This information may be quite rich, including not only the layout of the immediately perceptible environment, but also information about hidden objects, potentialities, affordances, and so on. Moreover, the underlying mechanisms that govern the dynamics of the system—which information is retained, and for how long—may be quite complex. In particular, the updating mechanisms may be entrained to environmental cycles. This allows animals to time their behaviour appropriately with respect to environmental processes, but without explicitly representing temporal information.

Although Hoerl and McCormack are less sanguine about denying that young children engage in temporal reasoning, they make a similar point about scripts. In order to successfully navigate a script, all an agent needs to do is to have the right action-sequences occur to them in the right order. The script can just be a programme run by the updating system, triggered by appropriate environmental cues. There is no need to for temporal relations between parts of the script to be explicitly represented. Of course the fact that children can provide at least partial verbal reports describing the structure of a script tells against this interpretation in this case. But insofar as nonhuman animals possess script knowledge, Hoerl and McCormack hypothesise that it is just implemented at the level of the updating system.

The picture is bold and parsimonious in the sparse cognitive resources it attributes to non-human animals. It remains to be seen whether it can be reconciled with the full

range and complexity of timing behaviours animals are capable of, or whether there are good empirical or theoretical reasons to take animals to be explicitly representing, and performing computations over, temporal information. I do not wish to take a stand on this complex psycho-semantic question. However, I do want to make the point that Hoerl and McCormack's and Campbell's respective pictures are not necessarily exhaustive of the options. It is possible to attribute to a creature *some* grasp of temporal notions, without thereby attributing a grasp of an overarching temporal domain to which those notions apply.

Suppose there are good grounds for maintaining that an animal explicitly represents, rather than being merely sensitive to, the amount of time elapsed since a significant environmental event. For example, it might be that the animal can combine and integrate this duration information with other types of information in a way which is too systematic and open-ended to be explained as the effect of interacting nonrepresentational oscillator mechanisms. We could then say that the animal locates the event in its immediate past, and thereby appreciates that, among other things, it is now unable to affect the occurrence of that event. (By contrast, Hoerl and McCormack would claim that all we need to say is that, at that time, the animal's updating system does not represent its environment as affording opportunities for whatever activities typically affect the occurrence of the event.) We might even have grounds for saying that the animal represents some event as having occurred earlier or later, or more or less recently, than some other recent event. This much could all be true without it possessing anything like a global conception of time—a domain of times on which it can plot episodes and events from distinct, temporally separated, stretches of experience.

What a thinker who uses just phase or script time is lacking is any understanding of how different instances of the same phase or script are related. Having this kind of understanding requires an ability, not only to orient one's ongoing actions and experiences in their immediate temporal surroundings, but also to integrate temporal information over distinct, separated stretches of experience. This means, at a given time, having a grasp of the relations between the world as it is now, and as it was or will be at times which are relatively remote from the present, and not related to it as parts of a script. Someone able to think in this way must have an understanding of the temporal world

as extending beyond their own immediate experience and activity. Moreover, as I argued above, it is only if a creature has this more connected conception of time that we can credit them with the counterfactual thought of things being otherwise from how they actually are.

We might therefore expect this kind of understanding of time to go together with a use of causal concepts that goes beyond what is involved in manipulating environmental variables. In the following section I explore the links between understanding of temporal and causal structure.

5.2 UNDERSTANDING CAUSAL STRUCTURE

Consider the interventionist approach to causal learning and inference discussed in chapter 3. The core of the concept of cause, on this approach, is if X is a cause of Y , then intervening on X should preserve the correlation between X and Y . This piece of information has immediate practical value for an agent who wants to affect Y but cannot directly do so: such an agent can affect Y by getting themselves into a position to intervene on X .

The notion of cause captured by the interventionist approach is a rather formal and schematic one, which has at its heart the idea of causation as a kind of difference-making—rather than, say, transmission of force or energy. There is no requirement of a causal mechanism, for instance involving spatial contact or energy conservation. It thus contrasts with accounts of the nature of causation, such as Salmon (1984) and Dowe (2000), and of causal cognition, such as Shultz 1982, which give a prominent role to the idea of physical process. Strikingly, there is not even any essential temporal component such as the requirement that causes precede effects. In interventionist accounts of causal learning, these additional requirements might be postulated as additional ‘substantive assumptions’ or learning heuristics, which aid or constrain agents in gathering causal information.⁶ But they are not at the core of the concept of cause itself. In one sense its minimality is part of the power of the interventionist account, since it points to ways that agents might discover asymmetric causal structure—not just the strength

⁶Cf. Gopnik, Glymour et al. 2004, p. 6.

or weakness of causal connections, but also what causes what—without relying on prior domain-specific assumptions about which kinds of things are causes and which are effects. But this also raises the question whether a thin, interventionist understanding of cause captures everything we associate with the concept, and in particular with the way in which a grasp of causal structure is connected with an thinker's conception of time.

Having noted this point, we can see how the local notions of pastness and futurity that go with a script-based understanding of time might be understood in interventionist terms. Where the agent has knowledge of causal relations between parts of a script, we can think of the script as just a special kind of causal graph, one representing a dynamic process whose phases are the variables that correspond to nodes on the graph. The graph represents the causal relations between the phases of the process.

An agent who thinks of parts of the script as within its power to affect will do so through its grasp of the possible interventions it can make on the script, and the typical effects of those interventions. For instance, it can affect some event *E* in the script by intervening either directly on *E*, or on its ancestors. Given that its ancestors all (temporally) precede *E* in the script, this means that, once *E* is past, there is no longer any opportunity for affecting *E*—until a new instance of the script occurs. The sense in which the past is fixed, locally to the script, is thus essentially a matter of timing one's interventions to occur at the appropriate point in a process.⁷

For a script-using creature, we can mark a distinction of sorts between type- and token-level causal beliefs. When such a creature locates itself in a given script, it has specific beliefs about what is causing what on that particular occasion, as well as standing beliefs about what generally causes what. For example, a child might know that, if they ask for juice, then juice will be served, and also believe that on this particular occasion no juice was requested and so none will be served. The combination of these beliefs, if they have any claim to being genuinely causal, requires the child also under-

⁷There is a delicate question in the background here about what the grounds are for attributing to a creature a representational understanding of the temporal and causal dimensions of a script they possess. Hoerl and McCormack's more deflationary view is that all we need to account for the timing behaviour of animals and young children is an association between environmental cues and types of activity, so that recognising a certain script, or phase of a script, prompts a learned behavioural routine. By contrast, I want to allow that an agent may be credited with some ability explicitly to represent temporal and causal relations between parts of a script, yet still lack the concept of a fixed past.

stand that, had they asked for juice, juice would have been served. However, as emphasised earlier, this thought need not amount to a genuine counterfactual. The child may not distinguish between the possibility of asking for juice on another occasion, and the possibility of having asked for juice on that very occasion. Rather, they may think of asking for juice as a generic possibility—where this notion of possibility is indiscriminating between something that might actually happen, and something that could have happened, but didn't.

A purely script-based thinker therefore cannot engage in historical causal thinking. Such thinking involves a contrast between the actual course of events and counterfactual alternatives; and I have argued that this contrast is not available to a creature who lacks a sense of the actual world as extending beyond the confines of a script. So how might a more connected conception of time be supported by a grasp of causality?

Both Campbell and Hoerl and McCormack, in work on temporal and causal cognition, have argued that the ability to think about time in a mature way is connected with a more sophisticated understanding of cause. Hoerl and McCormack, for their part, stress the ability to reason about the causal significance of temporal order as the crucial factor.⁸

An example of the kind of temporal-causal reasoning in question is provided by an experiment carried out by Teresa McCormack and Kerry McColgan (2008). In this study, children were shown a doll named Molly visiting a toy zoo. There were various animals that Molly visited on her trip around the zoo, and the test subjects (children 3–5 years old) were made to understand that she could only visit them in a certain fixed order. There were two version of the task: a search version and a planning version. In the search version, Molly was seen taking a photograph of one the animals (the kangaroo) with a Polaroid camera, and subsequently 'lost' the camera at one of the remaining animal locations. Subjects were then asked where the camera might be, knowing that Molly had used it to photograph the kangaroo. Only answers of locations which came after the kangaroo in the order of being visited were counted as correct. In the planning version of the task, subjects were told that Molly wanted to take a picture of the kangaroo, and had to decide where to put the camera so that she would be able to re-

⁸E.g. Hoerl 2008; Hoerl and McCormack 2011.

trieve it and take her photograph. In this version of the task, only locations before the kangaroo were counted as correct answers.

Both tasks require an ability to reason about the causal significance of temporal order. What is required here is a kind of joined-up thinking about place, time, and causality: in both version of the task, the children have to understand that what happens to Molly at an earlier time constrains what condition she is in, and hence what she can do, at a later time. For instance, Molly cannot take a photograph of the kangaroo if she has lost her camera at a prior location (search task), or if she has not retrieved it at a prior location (planning task). Secondly, they have to combine this with knowledge of the locations of the different animals, and the order in which Molly visits them—in the search task, to draw an inference about the possible location of the camera, in the planning task to make a decision about where to place it. The results indicate that the ability to think in this interconnected way emerges around 5 years of age; in general, the 5-year olds in the study, but not 3- and 4-year olds, were able to give correct responses.⁹ Other studies suggests similar developmental conclusions about the onset of causal-temporal reasoning abilities (e.g. McCormack and Hoerl 2005, 2007; Povinelli et al. 1999.)

The crucial point about the task is that it involves reasoning about a novel scenario. A test subject who was able to attempt the task repeatedly could work out just by trial and error that leaving the camera at location D does not cause Molly to take a photograph of the kangaroo, whereas leaving it at A or B does. Someone might acquire a script for the zoo scenario in this way that encoded causal relations between elements of the script. But this causal knowledge would be relatively minimal; in particular, it would not be grounded in a general appreciation of the causal significance of temporal order.

Hoerl and McCormack argue that this kind of causal thinking requires temporal reasoning, rather than mere updating, because it requires the agent to think through sequential steps in a temporal sequence to see how each depends causally on the previous one. For instance, Hoerl writes:

⁹In fact, in some simpler versions of the search task (involving just three locations), 4-year olds performed above chance. McColgan and McCormack suggest that this may be due to the fact that the simpler design made it possible for the 4-year olds to use the layout of the toy zoo as a spatialised representation of the order of events. If this hypothesis is right, then the study may not have been a pure test of temporal reasoning abilities, since the children effectively used the zoo as a visual aid.

...the causal relationship here is one where a particular outcome is dependent on a sequence of events, performed in the right order. As such, grasping it involves not just the ability to think about [one event followed by another], but also the child's seeing how [the desired outcome] depends on events actually happening in that order rather than any other. And this, in turn, allows us to credit her, not just with an ability to imagine [the desired outcome], but with the ability to think about [the outcome] as something that, even though it is not yet the case, will actually be the case in the future.¹⁰ (Hoerl 2008, pp. 497–498)

The child's grasp of the causal sequence of her actions thus enables her to think about the result—a novel, non-present state of affairs—as actual in the future.

However, I think this explanation does not quite get at the crucial issue. The question is how an agent might come to have a conception of time as a structure that extends beyond the horizon of their present activity. It is true that in order to solve either of the zoo task, the child has to be able to think backwards and think ahead to actual but non-present states of the world. But all this requires is that they be able to come up with a novel script for the situation, one which they can then employ to make sense of which actions of theirs will lead to a desired result. And it is not immediately clear how coming up with or employing a novel script requires anything like the connected and global conception of time I have been stressing is involved in mature thinking about the past.¹¹

John Campbell has a different account of how this kind of causal reasoning task connects with a global conception of time. His account stresses the need to recognise something like mechanism as a component of causality—something that is missing from the more minimal conception of causes as difference-makers. He writes,

We think that causal influence is transmitted from place to place by the movement of objects. I light an oil heater in the garden and then find it's cold indoors,

¹⁰The particular task Hoerl is mentioning in this passage is one devised by Cristina Atance, where a child is given an ant costume which comes in two parts that have to be put in a specific order. The task tests similar causal-temporal reasoning abilities to those involved in the zoo task described above.

¹¹This is somewhat simplifying the dialectic. In the same paper Hoerl objects to Campbell's interpretation of script-using agents as reasoning about their temporal past and future. Part of his argument is that it would be strange for a script-using agent to be able to think forwards or back to their immediate future or past without it occurring to them to continue the process indefinitely. (Ibid., p. 494) But this is not really so strange: the script gives practical meaning to the temporal relations between the events it contains, but not to others: the script does not tell the agent how to go on. This is connected with the fact, which I discuss below, that the causal relations encoded by a script are fairly narrow and stereotyped.

so I bring the heater in. Because of the movement of the heater, my intervention, my lighting of the gas at one place, makes a difference to the temperature at another place. We would be baffled by the idea that my lighting the heater outside could have made a difference to the temperature of the room without the heater having been moved, unless of course we appeal to some other story about heat waves or the motion of molecules. The movement of the physical thing, the oil heater, explains how it is that my intervention outside is making a difference to the temperature inside. (2006, p. 10)

Applying this to the toy zoo study, the idea is that kind of appreciation of how temporal order constrains what can happen at different locations is secured by a grasp of the causal properties of physical objects—in particular their ability to move sequentially through a series of spatial locations, and their tendency to stay put when left alone. For instance, in the search task, the spatial location of the camera at the end of the zoo visit is the causal upshot of it having been transported and then left there by Molly. This has consequences for what kind of state Molly was in (i.e. having the camera or not having it) at each stage of her journey, which in turn has consequences for what she could or could not do at each stage. Given this, plus the information that she actually took a photograph of the kangaroo, it is possible to infer a range of possible locations for the camera. Conversely, in the planning task, Molly's state at any given time of being with or without her camera is understood to depend causally on whether there has occurred a prior event of her retrieving it. This in turn means that whether or not she is able to photograph the kangaroo depends jointly on the order in which she visits the different locations, and the prior location of the camera. Campbell's point is that it is only because we understand ordinary physical motion as a causal process that we can understand, in a systematic way, how what happens at one place can depend upon what happens at another place.

Campbell further argues that this mechanism-based understanding of causality requires more than is provided by a phase- or script-based conception of time:

Suppose we have an intelligent agent operating with phase time only, making no distinction between one morning and another. An object may be observed one morning to be G when the previous morning it had been observed to be F. And our agent can recognize a counterfactual dependence of the G-ness on the F-ness of that object. If there had been an intervention on the F-ness of the object, it would not be G...formulating this kind of counterfactual does not seem to require use of

temporal ideas. What our agent cannot do is identify the ground of the counterfactual dependence. For the ground of the counterfactual is the persisting object itself, transmitting causal influence from its earlier place to its present place. Recognizing that temporal relation demands that our agent go beyond the confines of phase time, towards something more like our ordinary thinking about time. (Campbell 2006, pp. 10–11)

I do not want to rest too much weight on the idea of there being a notion of mechanism involved here that cannot be analysed in interventionist terms.¹² The point Campbell's remarks bring out which I want to underline is that recognising physical objects as sites of causal influence immediately raises the possibility of causal connections that cut across different instances of a phase or script. Doing something to an object at one time leaves a 'mark' on it that is transmitted to other places that the object visits, and affects how the object can interact with its environment at those other places. Once a thinker is equipped to think about causality in this way, there is no longer any reason for their recognition of causal connections to be confined to a single instance of a script.

The more general idea here is a script- or phase-based understanding of time tends to go with a relatively limited and unsystematic grasp of causality. Inter-script causal relations will tend to be restricted to a relatively narrow and stereotyped range of variables. Campbell calls this a 'phrasebook' understanding of causality, as opposed to a systematic one, explaining the distinction with a vivid example:

Suppose, for example, that you are at the bedside of someone in hospital, wired up to an array of tubes. The doctor is called away. Before he goes, the doctor hands you a gadget with a plunger on it, and says, "If he calls out in pain, depress the plunger, and do that once every couple of minutes till he seems comfortable." Here you certainly have some grasp of the use of the tool, but it is only a phrasebook understanding. Move you out of that context an inch and you have no idea what

¹²Cf. Woodward 2007, pp. 24–25: 'adherents of an interventionist account can readily acknowledge that information about causal mechanisms, properly understood, plays an important role in human causal learning and understanding. However, rather than trying to explicate the notion of a causal mechanism in terms of notions like force, energy, or generative transmission, interventionists will instead appeal to interventionist counterfactuals...information about a mechanism connecting C to E will typically be information about a set of dependency relationships, specified by interventionist counterfactuals, connecting C and E to intermediate variables and the intermediate variables to one another, perhaps structured in a characteristic spatiotemporal pattern.' Presumably the relevant interventionist counterfactuals will relate an object's exemplifying some property at one time with its still exemplifying it at a later time; and an object's travelling in a direction at a certain speed at one place and time with its presence in another place at a later time.

to do. Suppose you hit the plunger and an alarm sounds, what now? Suppose the patient seems delirious rather than in pain, should you use the plunger, or is that the one thing you should not do? Or would it be a better idea to rewire the plunger to hook it up to a different set of tubes? (2011, p. 175)

By contrast, a more systematic causal understanding will involve having a grasp of the underlying properties of the plunger as endowing it with a range of different causal powers that vary systematically over a range of circumstances, in something like the manner of Shoemaker's (2003) influential theory of properties.¹³ Without this deeper causal understanding, you may be completely at sea as soon as there is some slight deviation from the context in which you know how to use the object—from your script for using it.¹⁴

As I have said, there is plausibly a link between the more flexible and open-ended understanding of causality provided by an understanding of the properties of objects that underlie their causal powers, and being able to think beyond the temporal confines of a script. The former exerts pressure towards the latter, insofar as recognising the great variety of causal relations that might obtain between events means there is unlikely to be any obstacle in principle to recognising lines of causal influence that cut across local temporal contexts.

However, I think we should allow that these two dimensions of causal thinking are in principle separable. We can imagine an agent who is able to think about causal relations in a highly flexible way, constantly responding to novel problem-solving situations, but who lacks any use for an overarching temporal framework in which causal relations

¹³Peacocke 2011 argues for a similar claim regarding the importance of underlying explanatory properties in causal understanding. Peacocke argues for the somewhat stronger claim that a creature only counts as representing genuinely causal information if it can be credited with a grasp of the idea that the causal relation holds in virtue of an underlying explanatory property. While I am not unsympathetic to this position, I do not want to insist that there is no sense in which a creature with a grasp of the relevant interventionist conditionals is thinking causally.

¹⁴Campbell's distinction is somewhat similar to one Woodward (2007, 2011) draws between an 'ego-centric' viewpoint and a 'fully causal' viewpoint. An agent with a fully causal viewpoint "grasps that the same relationship that the agent exploits in intervening also can be present both when other agents intervene and in nature even when no other agents are involved." (2007, p. 32) (He also distinguishes an intermediate level, the 'agent causal' viewpoint, where other agents' actions are understood to exploit the same causal relationships as the agents' own.) However, the distinction is not quite the same, as an agent could have a fully causal understanding in Woodward's set that was rigidly limited to a narrow set of dependencies, and so not systematic in the sense Campbell is interested in.

between remote events can be plotted. Such an agent would retain only general information from each problem it solved, about the causal properties of kinds of objects and events and so on, simply discarding any information about the temporal and causal relations of the particular events involved in the last problem it solved. This agent's use of tensed notions would still be local to a given temporal context; the only difference is that the temporal context is not defined by a preexisting script, but rather is adapted to the demands of the unique causal situation. Such an agent would be constantly thinking up new scripts—devising novel motor routines for the situation at hand, guided by its grasp of general causal relationships that bear on the matter at hand—but its use of the concepts of past and future, fixity and openness, might yet always remain local to the latest script. Someone who thought like this would have no proper concept of the past. Despite all their causal sophistication, we could not credit them with a grasp of the temporal world as extending beyond the bounds of their present task.

There seems to be something strange about this problem-solving agent. They have the conceptual tools to trace causal relations between different problems they encounter and form causal hypotheses about where each problem came from and how they came to be working on it. Yet they have no interest in the past beyond whatever has a bearing on the present state of the problem. What would such an agent be missing?

5.3 MEMORY, AGENCY, AND CAUSAL SELF-LOCATION

I think the strangeness of a purely problem-solving agent lies in the fact that it has no interest in locating itself as an element of the world that it acts in. It fails to realise that it itself is the common element tying together the different problems it faces. This suggests that there is a close connection between a proper concept of the past, and a conception of oneself as an item in the causal order. Such an understanding will involve an appreciation, not just of the causal powers of the objects one interacts with, but further of the ways in which one is both shaped by, and makes a difference to, the world in which one lives. M. G. F. Martin sums up this point in relation to episodic memory:

We can then think of the plight of the infant in coming to have a full grasp of the notion of the past...as taking the following form. The infant needs to make sense of how there can be specific, and hence actual, events of which it has know-

ledge or conscious awareness, but which are nevertheless not part of the present scene. We can point to two things which might help it in this predicament: first a grasp of a concept of time as a causal structure in which earlier events cause later ones; secondly an understanding of itself as located within that causal structure, such that recalled experiences have to possess a causal location prior to that of the present episode of recall. Past events are no less actual than current ones, but they are located in a different part of the structure that is time. (2001, p. 280)

Martin's point is that a full grasp of the notion of the past is connected with a capacity for episodic recall via the memorial subject's understanding of memory as a causal process: a process in which particular experiential episodes leave their mark on the subject and thereby remain available for recall at a later time.

Martin is clearly right to stress the central role that a grasp of the causal character of recall has in a conception of oneself as an element of the causal order. But to this I think we should add that our ordinary conception of the past, and of ourselves as extending into the past, includes more than passive receptivity: it also includes the idea that one has, in however small a way, acted on, and left one's mark on, the world which one now experiences and acts in. This is, perhaps, not a necessary truth about any intelligible conception of the past, but rather one about what our actual conception of the past is like, given our nature as active beings. We are not merely observers and repositories of memorial information: we are full participants in the causal order, and leave our marks on the outside world no less than we are impinged on by it. Our ordinary thinking about the past reflects this fact.

Stressing the importance of one's past agency is by no means in tension with giving a central epistemological role to memory in our thinking about the past. Indeed, given our active nature, these two features are mutually supportive. One point is the importance and ubiquity of agential memory: memory involves not just passive perceptual experience, but also experiences of doing things. Indeed most autobiographical memories are of experiences, in a broad sense, that fuse agential and perceptual components: one remembers the birdsong on a woodland spring walk, or a friend's facial expression when one gave the bad news. More broadly, though, one's grasp of the causal mechanisms by which memorial content is acquired and preserved is not separable from an understanding of one's past agency, insofar as this encompasses where one has been and what one has done. The kind of causality involved in memory is that of an extended experiential

route through the world. Given our active nature, that means, in part, an agential route through the world. One's life as an information-gatherer and one's life as an agent do not move along separate causal tracks. There is continuous causal commerce between them. Thus, the kind of causal self-location involved in locating one's memories in the past is, for creatures like us, interdependent with the project of locating one's agency as a small part of the history of the world.

The previous chapter discussed some of the practical value of regret, and in particular anticipated regret, as a way of making sense of how one's current motivations relate to one's value and longer-term interests. But in the context of the present discussion, we can see how agent-regret also serves an important cognitive function, in allowing one to recognise past actions and motivations as distinctively one's own, which one would nevertheless not presently endorse. Agent-regret has a dual character: it involves a kind of negative assessment, in which one distances oneself from the action and its motives; and at the same time a kind of ownership, and an understanding 'from the inside' of the motives that led to the action. In taking this kind of attitude to one's agency, one thereby recognises oneself as more than just the sum of one's present beliefs and desires, but rather as a being that undergoes, in Taylor's phrase, maturations and regressions.

What I want to propose is that this fairly thick, ethical understanding of oneself as persisting through maturations and regressions is not parasitic upon a prior grasp of a temporal framework upon which one's maturations and regressions can be plotted; rather, an understanding of one's maturations and regressions is part of what supports one's grip on the temporal framework. As agents who are conflicted and changeable, our persistence over time involves persistence through changes in motivational and affective states. One needs to grasp how one could remain the same agent over time through changes, even radical changes, in what one is doing and what one wants. It is not to the point to appeal to bare spatiotemporal and causal continuity here, since part of what is in question here is why one should apply spatiotemporal and causal notions beyond the confines of a script-like temporal-causal map of a particular task—why one bothers to identify oneself as a common element to the different problem situations in which one finds oneself.

Part of what emotions like regret do for us, I suggest, is articulate a recognition

of one's continuity through changes in one's practical situation, and thereby support the identification of an agential route through the world. This can happen on a fairly small scale: a child eats a whole box of plums, then gets a tummyache and regrets it. One might think of the regret just as a signal to avoid eating whole boxes of plums in the future. But we can also discern something more complex. Insofar as the regret is attached specifically to the memory of eating the plums, we can see it as involving a recognition of the transience of the past desire for more plums, and so as supporting the child's identification of itself as persisting through its transient desires. These structures of affect and self-recognition thus provide some of the conceptual tools for an agent to see itself as extended in time, and so to acquire a use for a global temporal framework in which particular past events can be distinguished and plotted.

Let me take stock of the argument of the last three chapters. I have been arguing, along various converging lines, for the importance of a certain type of causal thinking, historical causal explanations of particular events. This notion of cause is distinctive in the way it relates to the actual world: it involves explaining, of some actual event, why that event occurred. The interest we have in explaining actual events in this way goes beyond any interest in finding out what, in general, causes what, in order better prepare to ourselves for similar causal situations we may encounter in the future. One mark of this, as I noted in chapter 3, is that historical causal judgments are insensitive to the presence of pre-empting and overdetermining factors. We are interested ultimately in tracing the actual path of causal influence, rather than the distribution of probable outcomes in relevantly similar possible situations.

In the previous two chapters I argued that part of being a self-conscious agent is the way in which one applies these forms of historical causal thinking to oneself, locating one's past actions and interactions in a wider nexus. What this chapter has attempted to bring out is the way in which this kind of causal self-location is connected with an ability to think objectively about time. There is support running in both directions: on the one hand, applying historical causal thinking requires a distinction between the actual and the non-actual, which needs to be articulated in terms of a the actual world as a global temporal framework. On the other hand, possession of a global temporal framework

is, at least for us, supported partly by the recognition of oneself as a common element to the various situations and tasks one encounters, as a basis for tracing causal relations across different local temporal concepts.

The crucial point is that causal self-location is associated with ways of thinking about oneself that one has only in relation to the past. This is partly because of the critical role played by episodic memories of one's own actions. More generally it is because, as I have argued in the previous two chapters, one's temporal orientation as an agent imposes a limit on what one can know about the causality of one's present and future actions. The kind of reflective assessment involved in agent-regret exemplifies a distinctive insight into one's causal structure: an insight into the maturations and regressions that make up one's life as an agent. By contrast, knowledge of one's present and future causality is mediated by one's practical knowledge of what one is doing, and so is partial—both in the sense of being incomplete, and in the sense of being interested.

What this means is that our understanding of the global temporal framework as it extends into the future is, in a sense, ungrounded. One has of course a general grasp of the temporal framework as it extends into the future, and one can in principle plot temporal and causal relations between projected or hypothesised future events. Perhaps one's mastery of a calendar system can allow one to pick out, by definite description, actual particular future events, like the 2020 European Championship. What one lacks is the engaged but reflective knowledge of one's causality—of the particular actions and experiences that make up one's causal history—that would enable one to locate oneself in the future in the way one can locate oneself in the past. One cannot see oneself in the future in the way one can see oneself in the past. This asymmetry of causal self-location is, I suggest, key to understanding how we can think objectively about time without falling into fatalism.

This chapter has been concerned with the idea that locating events—temporal particulars—in the global framework is involved in thinking objectively about time. But I have tried to do this while remaining fairly neutral about what such particulars are, and what is involved in singling them out. I will take up these issues in the remaining two chapters.

6 | TENSE AND TEMPORAL PARTICULARITY

The last chapter argued that part of what is distinctive about ordinary thinking about the past is that it involves thought about temporal particulars. But I have not said much about what is involved in the idea of a temporal particular. On one way of thinking, a temporal particular is a particular moment in time, or perhaps a fact about something happening at a particular time. On this view, the identities of temporal particulars are dependent on the theoretical identification of times. A point going back to Kant's *Transcendental Aesthetic* is that we cannot directly perceive times, and so their individuation depends on a level of conceptuality that organises, or imposes order on, our understanding of time. One might take more or less realistic attitudes to the identification of particular times, but they remain, in some sense, theoretical entities: inferred, posited or constructed from non-particular raw materials.

Against this, one might think of temporal particulars as concrete events, which occupy time in something like the way that ordinary objects occupy space. This view is associated in particular with the work of Donald Davidson. Davidson 1980a On this view, the individuation of events does not fundamentally depend upon their differentiation within a theoretical framework. Rather, particular events are, by their nature, unrepeatable.

Someone who thought of temporal particularity as a theoretical construction might think that, in general, we single out particulars by employing a dating system that assigns a unique real number to each time. The dating system is the conceptual grid that organises temporal goings-on into a framework of particulars. I want to allow that we can single out events in this way, and that clock and calendar systems offer a singularly powerful and versatile means of constructing uniquely identifying descriptions of

events. But I will argue that this cannot be the sole or primary basis of our ability to think about temporal particulars. Rather, such command as we have of the temporal framework must rest on a prior capacity to single out concrete particular events.

This chapter will discuss in more detail what is involved in the idea that events are particulars. The main aim is twofold. First, the discussion will seek further to cement the idea, which emerged from the previous chapter, that ordinary temporal thinking, and in particular thinking about the past, involves a sensitivity to temporal particularity. Secondly, it will clarify the commitments this involves, and in particular will argue against the idea that this dimension of particularity cannot be adequately understood as an abstraction from temporally non-particular, tensed facts. The discussion thus involves a transition from temporal thought to metaphysics. In general, the idea that the structure of thought about a domain involves sensitivity to particularity is not metaphysically neutral with respect to the structure of the domain: it involves commitment to the idea that the domain contains structure enough to warrant describing it as containing particulars.

Section 6.1 is a brief introductory discussion of the relevant notion of particularity, and its connection with failure of the Identity of Indiscernibles. 6.2 discusses particulars and their individuation in connection with the idea of a spatial framework. 6.3 carries over the discussion to the temporal case, and rehearses the theme of the previous chapter that properly counterfactual thought requires locating particular events in a global temporal framework. I introduce the opposing position of someone who takes the tenses as primitive, and sees the temporal framework as fundamentally an abstraction from tensed truths. The best-developed form of the tense primitivist's position is exemplified by Arthur Prior's philosophical use of tense logic to formulate theses on the structure of time. In section 6.4 I therefore consider the idea that tense logical axioms express postulates about the structure of time, in particular the topology of time. I argue that this rests on a subtly question-begging interpretation of the tense operators; it is only if we already presuppose a conception of time as a framework of particulars that we are warranted in interpreting the tense operators as expressing relations capable of generating an order. Thus I argue our knowledge of temporal particulars cannot be explained in terms of a primitive grasp of the tenses; rather, we also have to acknow-

ledge an element of intrinsic particularity in temporal experience, which supports our possession of a global temporal framework.

6.1 PARTICULARITY AND INDIVIDUATION

Before discussing specifically the temporal case, I will briefly sketch what I take to be the general issues at stake in the question whether a system need be understood to contain particulars. The notion of a particular, as I understand it here, is best distinguished in terms of a failure of a principle of the Identity of Indiscernibles. Here is a statement of the principle:

II: If a has all and only the qualitative characteristics of b , then $a = b$.

One might wonder how informative this condition is.¹ The restriction to ‘qualitative characteristics’ is necessary to avoid trivialising the principle. If the antecedent is formulated using a more general notion of a predicate—say, ‘everything true of a is true of b ’—predicates as ‘is identical with a ’ will be included, in which case the principle is trivially satisfied. But, on the other hand, the restriction to the qualitative seems like a restriction just to something’s shareable features; in other words, those features which are not relevant to its identity. So it looks as though failure of **II** fails to state a substantive condition for something’s being a particular, since its interpretation relies on prior intuitions about the which predicates are relevant to the identities of particulars.

This point does not, however, mean that failure of **II** is completely uninformative in clarifying the notion of a particular. The notion of the qualitative and the particular are correlative: intuitions about the identities of putative particulars will tend to be informed by intuitions about which predicates express qualitative characteristics, and vice versa. But it is distinctive of particulars that some such distinction is available: for any particular, we can distinguish questions of *what it is like*, from questions of *which one* it is. A different way of saying essentially the same thing is that there should be available a distinction between qualitative and numerical identity.

¹MacBride 2005, 609, n. 41 expresses scepticism that appeal to **II** is useful for formulating a particular/universal distinction.

By way of contrast, consider colours—not coloured things, or their particular colours, but colours themselves. Colours vary along dimensions of hue, brightness and saturation, and can be distinguished by specifying values for those dimensions. We might represent these as axes and then distinguish colours from their position in the colour solid. Once we have specified a colour in that way, there are no further questions to be asked about which colour it is. The question of which colour is exhausted by the question what it is like with respect to all ways in which colours can vary. There is no distinction of qualitative and numerical identity available.

Conversely, it part of the way we ordinarily think that there is no way in principle of specifying what a particular concrete object is like—say, a factory-made pencil—that will settle the question which thing it is. Even if it does turn out to have some mark which distinguishes it from all similar individuals, this is not implied by its distinctness. That there need not be any such specification available to ensure the individuality of a particular was denied by Leibniz:

...it is not true that two substances can resemble each other completely and differ only in number, and that what Saint Thomas asserts on this point about angels or intelligences (that here every individual is a lowest species) is true of all substances, provided that one takes the specific differences as geometers do with respect to their figures. (Leibniz 1991, p. 9)

On a Leibnizian view of the matter, there is a more fundamental level of description of reality—one available to God—on which particularity drops out, and is replaced by the general pattern of specific kinds and attributes (where ‘specific’ connotes a thing’s species, rather than its numerical identity.)

Perhaps there is in principle some description of the global pattern on which reference to the identities of particulars drops out, just as this same level of description makes no mention of cause.² However, for reasons discussed in previous chapters, I do not think this means we should conclude that particulars, or causal relationships are unreal, but rather that they are features of the world that can be discerned only from the perspective a thinker who occupies some local position. This does not make facts about particulars dependent upon, or relative to, a thinker’s perspective in any way that would undermine their claim to reality.

²See, for instance, Dasgupta 2009 for a contemporary Leibnizian theory along these lines.

The intuition that ordinary objects are not individuated by qualitative features alone is clearly connected with our ordinary understanding of spatiality. We do not think of spatial locations, or spatial relations, as qualitative features of objects. Rather, we think of how an object stands spatially as a separate matter to what it is like. Conversely, we think of space as a framework structured independently of the distribution of qualities through it, and providing an arena in which objects bearing those qualities may be located and thereby individuated. This is not to say that our ordinary conception of space is committed to anything like absolute position. The point is just that the spatial distribution of objects is not determined wholly by their relations of similarity and difference along various dimensions of qualitative variation.³ The intrinsic structure of the spatial framework provides an order, or system of relations, by means of which objects may be distinguished otherwise than by their qualitative variation.

This raises the question what our basis might be for taking space to have this character as a dimension of particularity, and what metaphysical commitments an adequate account of that basis might incur. Seeing what is involved in answering these questions will then allow us to carry the discussion over to the temporal case.

6.2 SPATIAL INDIVIDUATION

In an early discussion, Russell poses the question how spatiality is involved in the individuation of distinct instances of the same sensible quality. He considers the case of visually perceiving two patches of white, at different places in the visual field.

Let us suppose...that within one field of vision we perceive two separated patches of white on a ground of black. It may then be taken as quite certain that the two patches are two and not one. The question is: Can we maintain that there are two if what exists in each is the universal whiteness? (1956, p. 116)

Saying that they are distinguished by the respective spatial position is not very satisfactory, Russell thinks, since this just pushes the problem of individuation onto some

³Similarly, this is not to say anything inconsistent with the lesson of General Relativity that the affine and metrical structure of spacetime—and, in cases of wormholes and the like, the topology—is not independent of the distribution of energy. It is just the minimal idea that the features of the spatial (or spatiotemporal) framework relevant to individuating particulars—which include the topology, but not the affinity or metric—are not wholly determined by non-spatial structure.

more obscure entities, spatial positions. Moreover even if absolute spatial positions exist, they cannot be perceived directly, and so the distinctness of their spatial position cannot provide a basis for the judgment of distinctness.

In some cases, perhaps, the two patches might be distinguished with reference different qualities they are 'co-present' with. For instance, perhaps one patch is round and the other is square; since roundness and squareness are incompatible, they must be distinct. But,

It is obvious, however, that this method of distinguishing the two patches is altogether inadequate. The two patches are just as easily distinguished if both are square or both are round. So long as we can see both, no degree of likeness between them causes the slightest difficulty in perceiving that there are two of them. The difference of shape, whether it exists or not, is not what makes the patches two entities instead of one. (pp. 116–117)

He then considers the different possibility that such particulars are individuated by their relations to one another. He writes:

It may be said that the two patches are distinguished by the difference in their relations to other things. Suppose a surface of black with a small white space in the middle...Suppose we have another white patch, of exactly the same size and shape, entirely surrounded by red. Then, it may be said, the two patches of white are distinguished by difference of relation, since one is surrounded by black and the other by red. But if this ground of distinction is to be valid, we must know that it is impossible for one entity to be both wholly and immediately surrounded by black and wholly and immediately surrounded by red. I do not mean to deny that we do know this. But two things deserve notice—first, that it is not an analytic proposition; second, that it presupposes the numerical diversity of our two patches of white. (p. 117)

It might perhaps be replied that the relevant proposition *is* in fact analytic, if the 'wholly' in 'wholly and immediately surrounded by' is read as 'exclusively'. Setting this aside, Russell's main point is that the immediately perceptible spatial relation which one white patch bears to the surrounding black is not logically incompatible with the same thing's bearing the same relation to another, (qualitatively) distinct colour. Nevertheless, we do take ourselves to know that one single thing cannot bear this same relation to two distinct things—and, Russell claims, we know this on the basis of visual awareness. More generally, he says, we know of the spatial relations of perceived objects that,

They, or some of them, must be asymmetrical, *i.e.*, such that they are incompatible with their converses: for example, supposing “inside” to be one of them, a thing which is inside another must not also be outside it. They, or some of them, must also be transitive, *i.e.*, such that, for example, if x is inside y and y is inside z , then x is inside z —supposing, for the sake of illustration, “inside” to be among fundamental spatial relations...It follows that at least some of the fundamental spatial relations must be such as no entity can have to itself. It is indeed self-evident that spatial relations fulfil these conditions. But these conditions are not demonstrable by purely logical considerations: they are synthetic properties of perceived spatial relations. (pp. 117–118)

Russell thereby connects the individuation of particulars as numerically distinct with the construction of a system of spatial relations which satisfy the relevant ordering properties.⁴ Understanding the question of the individuation of spatial particulars in this way, he claims, is an improvement on framing it in more intuitive terms:

The essential characteristic of particulars, as they appear in perceived space, is that they cannot be in two places at once. But this is an unsatisfactory way of stating the matter, owing to the doubt as to what a “place” is. The more correct statement is that certain perceptible spatial relations imply diversity of their terms; for example, if x is above y , x and y might be different entities. (p. 121)

There is room for doubt about how closely these two cognitive projects should be tied together; that of distinguishing spatial particulars, and that of constructing an explicit intuitive geometry. Indeed, Russell’s so connecting to some extent goes against his insistence that the numerical distinctness of the particular white patches is just immediately apparent: ‘it is self-evident that, in such a case, there are two different patches of white.’ (pp. 112–113)

To bring this out, consider the fact that the relation of *being inside* is in fact not necessarily either asymmetric or irreflexive: for instance, neither property holds on the surface of a torus. Similarly, neither are relations of *left* or *above* on the surface of a sphere. (Where irreflexivity and asymmetry fail, familiar two-place relations are not

⁴There is an unclarity in Russell’s presentation. The idea that particulars are individuated by their standing in mutual ordering relations is different from, and not obviously connected to, the idea canvassed in the previous quotation that particulars might be distinguished by their having incompatible relational properties, like ‘being wholly and immediately surrounded by red’ and ‘...by black’. From Russell’s surrounding discussion it is clear that he takes the former idea, of a system of transitive, irreflexive and asymmetric relations, to be the more fundamental.

sufficient to express order, and a four-place relation is required instead.) We can imagine discovering that our world in fact has a non-standard topology, so that the ordering properties of spatial relations are different from what we had supposed. But this would have very little bearing on our ability to make judgments of distinctness on the basis of spatial perception.

The obvious reply on Russell's behalf is that he is here concerned with the features of perceptible, not physical, space, and whilst the topology of physical space may be a matter for scientific investigation, the geometry of perceptible space is immediately obvious. Against this, we might note that it is easy to imagine situations in which only very limited investigation revealed one's immediate spatial environment to have an unusual topology—for example, I walk forward in a straight line, and after a short time one confront exactly the same visual scene as I confronted when I started.⁵ In such a scenario, it might become far less obvious that the spatial relations between perceived items had the relevant logical properties, even at a single moment of perception in which nothing moved.

Of course it would be open to Russell to insist that the visual field, narrowly construed as the array of visual sense-data at a single moment, does necessarily and manifestly have a certain intrinsic geometry, and that this is invariant with respect to the wider context of environmental spatial awareness in which a given visual experience is embedded. But whether or not this is correct is a complicated, and perhaps empirical, further question. On the other hand, we might do better simply to disentangle questions of topological structure from the question how perception presents us with sensible particulars. It would be open one simply to note, as indeed Russell does earlier in the same address, that perception sometimes presents us with distinct items *as* distinct, and to note this without committing to the idea that it does so by presenting us with the materials for a consistent and universal geometry of perceptible space.

The deeper point here, half-acknowledged by Russell but obscured by his subsequent remarks about the logic of spatial relations, is that we can only even begin to theorise about the ordering properties of a system of relations—properties like irreflexivity, whose definition involves identity—once we have already provided for a distinction

⁵Many such scenarios are discussed in Reichenbach 1956's rich discussion of visualising non-Euclidean space.

between numerical and qualitative identity for the elements of the domain. In the spatial case these may be the dimensionless points of analytic geometry, or Russell's sense-data, or ordinary objects. Whatever they are, questions such as whether the space in which they exist is topologically closed, or merely repetitive, only get off the ground once we can distinguish between qualitative and numerical identity for these items.

It might be thought that, once it is established that we are dealing with a relation, it will always be in order to ask whether the relata are identical or not. If a is to the left of b , surely it must make sense to ask whether a is identical with b , and thereby to begin to gather data bearing on the question of the ordering properties of the relation 'to the left of'. This thought is half right. It is right in that, if we have a relation in the sense of something perspicuously expressed with the relational statement $R(a, b)$, then we can intelligibly enquire into the identity of a and b . But to assume that relationality must be expressed this way is just to assume a and b are particulars—that is, that relations relate particulars.

That this assumption is not required can be seen by considering the position of someone who rejects the category of particulars. Such a person holds that all facts about particulars are reducible to ones concerning universals.⁶ Someone who holds this universalist position will prefer to express relational facts in terms of the 'compresence' of a complex of monadic and polyadic universals. On this picture, a relation is just a polyadic universal. For instance, what a defender of particulars would describe as an F-thing standing in relation R to another F-thing, the universalist might express with the statement, ' R is compresent with $(F;F)$ ', where F is something repeatable.⁷ But this universalist formula fails to distinguish between the cases where the F-things are distinct and where a single F-things bears R to itself. The particularist will accuse the universalist of running together distinct possibilities; the universalist will claim that the particularist spuriously multiplies them.⁸

⁶Russell in *Human Knowledge* holds such a theory; for a more recent defence of a 'bundle theory' of particulars see Hawthorne and Cover 1998.

⁷This kind of expression is used in Hawthorne and Sider 2002.

⁸N.B. the universalist may be able to employ some further ideology in order to distinguish these possibilities; for a detailed discussion of the options see Hawthorne and Sider, *op. cit.* The present point is just that the universalist's ideology gives us a way of representing relations without necessarily thereby making any distinctions corresponding to those that the particularist makes in terms of the identity or distinctness of particulars.

This is a point about a possible metaphysical disagreement. But it has a conceptual or epistemological reflection; namely, that someone operating without the category of particular, and without the correlative distinction of qualitative and numerical identity can nevertheless employ concepts that are relational, at least in the sense of being poly-adic. Someone who thinks only in terms acceptable to the universalist will be able to think that whiteness is to the left of whiteness, for instance, without thereby inviting the question whether the whitenesses so related are identical or distinct. The force of this point is that we can, at least in principle, discern relations between things without supposing that the relata are particulars, and hence without conceiving of the relation as being capable of generating an order.

A different way of putting Russell's point, then, would be to say that, at least sometimes, spatial experience provides adequate grounds for judgments of numerical distinctness, independently of apparent qualitative variation. But, contrary to what Russell says, this need not necessarily be explained in terms of a synthetic *a priori* apprehension of spatial order. Rather, we might instead see the capacity spatially to individuate particulars on the basis of perception as prior to, and supporting, the ability to reason about order. This is not to say that perceptual spatial judgements of numerical distinctness are not fallible or defeasible. Double vision is of course common; and we can perhaps imagine other cases in which a perceiver discovers themselves to be in an unusual spatial or optical environment, and thereby revises their initial judgment of distinctness, which do not obviously involve illusion.⁹ The point is rather that spatial perception, in particular vision, is intrinsically such as to provide the perceiver with perceptual access to particulars as such, to the effect that questions of their numerical identity and distinctness may be at least meaningfully raised, and frequently settled, on the basis of such perception.

The intrinsic particularity of visual perception is expressed by M. G. F. Martin:

When I look at a duck in front of me, I am not merely presented with the fact that there is at least one duck in the area, rather I seem to be presented with *this*

⁹One obvious example is mirrors—e.g. Steenhagen 2017 argues that mirror reflections are not illusions. More fancifully, the kinds of adventures in locally non-Euclidean environments described by Reichenbach (op. cit.) furnish examples where further exploration and experiment by the observer might lead to revision of their initial identity judgments.

thing (as one might put it from my perspective) in front of me, which looks to me to be a duck. Furthermore, such a perception would seem to put me in a position not merely to make the existential judgment that there is some duck or other present, but rather to make a singular, demonstrative judgment, that that is a duck. My grounds for an existential judgment in this case derives from my apprehension of the demonstrative thought and not vice versa. (2002, p. 173, original emphasis. Cf. also Soteriou 2000.)

Later, Martin gives a more theoretical elaboration of this thought as follows:

Once we reflect on the way in which an experience has a subject matter, the presentation of a particular scene, then we need a way of making room for the essentially or inherently particular aspects of this as well as the general attributes of experience. We need to contrast the unrepeatable aspect of its phenomenology, what we might call its *phenomenal nature*, with that it has in common with qualitatively the same experiential events, what we might call its *phenomenal character*.

The important points here for my purposes are, first, that perceptual experience, specifically visual experience, puts its subject in a position to think singular thoughts about perceived objects, so that the subject can then raise such questions as ‘is *this* the same one as *that*?’ By contrast, an experience that supported only a pair of existential judgments, ‘Here’s an F’ and ‘There’s an F’ would not allow for this question to be raised without further descriptive complexity (‘is the first F the same as the second F?’) Secondly, the way in which visual experience puts the subject in a position to think such thoughts is due to its intrinsic nature as an episode in which the subject is presented with some particular scene and its components. The particular perceived objects figure in the perceptual situation not just as possible values of bound variables in the perception’s content; they themselves are present in the subject’s experience.¹⁰

¹⁰I am not certain if this reading is exactly what Martin intends; in the context of the quoted essay, his official goal is to argue that an intentionalist account of the contents of experience is consistent with both allowing for experience’s intrinsic particularity, and giving a uniform account of perception’s intentional contents in cases of veridical perception and hallucination. Given what else he says, it is not clear to me whether he would want to endorse the claim that it is the particular phenomenal nature of the experience that puts the subject in a position to think singular demonstrative thoughts about its subject-matter, rather than the fact that the experience has as its intentional content a function from contexts to propositions which determines, in that context, a certain singular proposition. However, the present context provided by Russell’s question about the basis of judgments of numerical identity and distinctness provides some further motivation for the idea that capacities for singular demonstrative reference are grounded in the intrinsic particularity of perceptual experience, rather than their intentional content.

This kind of story about the particularity of perception brings out starkly a theme that has come up several times now: in giving an account of how a subject is in a position to make judgments of numerical identity and difference amongst the objects they perceive, we cannot remain neutral about the actual structure of the environment in which the subject is embedded. Specifically, we are committed to the idea that the subject's environment contains particulars of which the subject is aware. Furthermore, it does so in a way that requires us to draw a distinction between qualitative and numerical identity for those objects. For the intrinsic particularity of the subject's perceptions consists just in the fact that they are viewing one particular object rather than another. Even if, as a matter of fact, there is always some qualitative mark distinguishing the various objects in their environment, this qualitative difference is just irrelevant to the experience's intrinsic particularity.

On the other hand, we should also recognise that the intrinsic particularity of experience alone is not in general a sufficient basis for much substantive knowledge of particulars. In order for a subject to make competent and knowledgeable judgments of numerical identity and difference among the objects of their perception, they plausibly must employ further conceptual capacities, including at the very least a basic understanding of the spatial structure of their environment.¹¹ But, in recognising this, we need not take the quasi-Kantian position that Russell's discussion leans towards at points, that the particularity of spatial experience is the product of imposing a kind of conceptual grid to order its contents. Rather, we can see knowledge of particulars as the joint upshot of the intrinsic, nonconceptual particularity of individual experiences and one's general, fallible and revisable conceptual command of the environmental spatial framework in which those experiences are embedded.

With these general remarks on particulars and their individuation in mind, we can consider how the same issues play out in the temporal case.

¹¹Perhaps much more is required. Perhaps, as Wiggins 2001 argues, singling out particular often requires conceptual capacities associated with knowledge of that thing's kind. This is entirely consistent with the point about the intrinsic particularity of experience.

6.3 TENSE AND TEMPORAL PARTICULARITY

In the the previous chapter I argued that our concept of the past as fixed requires a global temporal framework in which particular events and their relations can be plotted. It is only in the context of the temporal framework that we can draw the contrast between the actual and the counterfactual presupposed by mature thinking about the past. And this in turn requires a flexible and open-ended grasp of temporal relations as capable of generating an indefinitely extendible order, not limited to the confines of a script or phase.

Our discussion of Russell would suggest that possession of the global temporal framework is intimately connected with the ability to single out particular events. It is only if one is given some particular that one can entertain questions of numerical identity or distinctness with other given particulars. Possession of a framework within which particulars can be located is interdependent with, rather than prior to, this ability. Relatedly, one need not have comprehensive or accurate knowledge of the actual topological properties of the framework in order to be able to single out a particular within it. One can entertain the possibility that one is in a cyclical structure, and is encountering the same thing again, rather than some new thing, without losing one's grip on the particular thing singled out. Indeed, as I have been pressing, these questions can only come into view given a more basic grip on the particulars that populate the domain.

In the temporal case, we normally think that time is fully linear, and that temporal particulars, once encountered, cannot be re-encountered. But in view of the above discussion, this is perhaps not strictly necessary for the kind of grasp of the temporal framework that I am suggesting. Perhaps time is not in fact a linear order. In Special Relativity, for example, timelike separation is a partial order. Of course one might think that a mature understanding of time involves the idea that at least one's immediate local timeline is fully linear.¹² But we can equally imagine someone who takes seriously the possibility of local time travel and causal loops, without thereby giving up their conception of time as a framework of particular events. This is not to say that ordinary thinking about time

¹²Campbell 1997 for instance argues that an understanding of the causal structure of episodic memory requires an understanding of one's personal timeline as linear. Perhaps that is right, but the main point I am making here does not require it.

does not involve assumptions about the topology. The hypothesis of cyclical time certainly encourages fatalist intuitions, and puts pressure on our ordinary concept of the past.¹³ But any such knowledge we have of the overall structure of time should be seen to be supported by, rather than a precondition of, our being given temporal particulars.

But someone might insist that our understanding of time is fundamentally not like this. It might be thought, rather, that the concept of something's being past is just a primitive and irreducible matter, with the connotation that what is past is fixed and inaccessible to intervention. The difference between its raining and its having rained is a difference in kind of information, rather than a difference in the location of a particular within a framework.

Someone who holds this view would likewise object to the idea that locating something in the past is a matter of discerning a temporal relation between particulars. The idea that spatial experience presents us with particulars is fairly intuitive. By contrast, it is far less clear that experience possesses any analogous temporal dimension of particularity. Rather, it may seem that experience just presents us with things happening, rather than particulars happenings.

The challenge for this view is to explain how the primitive concept of past fixity differs from one employed by a creature who thinks only relative to a script. For a script-using creature, pastness has an immediate practical meaning, reflected in the fact that, when a phase of the script is located in the past, it is no longer possible to intervene. I argued that a script-using creature does not properly represent the past as fixed, because they do not have the resources to distinguish the possibility of intervening in a future instance of the script from the counterfactual possibility of having intervened in the instance just past. It is only if an event is understood as a one-off that its current inaccessibility can amount to fixity as we understand it.

We can put the point more vividly. Think of a script-using creature as sorting information about its temporal surroundings into three boxes: past, present, and future. Information in the 'past' box concerns things that can no longer be affected or interrupted; the 'present' box concerns things to be acted on immediately, and 'future' things to be planned for. The boxes are effectively an out-tray, an urgent in-tray and a non-

¹³For a rich discussion of closed time and its ethical implications, see Sorabji 1988, ch. 10.

urgent in-tray, respectively. As time goes on and the creature gets more information, the creature *updates* the contents of the boxes. Some information is retained, other information is discarded or overwritten. Information which is in the 'past' box at one time may turn up in the 'future' box at another time. There is thus no sense in which placing something in the 'past' box means regarding it as fixed, since its being placed there does not rule out its being placed somewhere else in future. If the contents of the boxes are continually updated, then nothing is fixed.

What the script-using creature's use of the tenses is missing is, I have argued, the idea that tense locates events in a temporal framework, rather than merely categorising them. But the primitive tense theorist holds that being in the past is a basic and irreducible kind of temporal information. The challenge is thus for them to explain how our use of the tenses differs from one in which tense is just a label that can be updated, discarded or overwritten. In particular, they need to explain why an event's being categorised as past means it must remain that way.

One may respond to this challenge by saying that our use of the tenses is different from that of a script-using creature because it enjoys a greater degree of systematicity, but this systematicity need not be explained in terms of a grasp of temporal relations between particulars. For instance, we understand that if something is past, it will always be past, whereas a creature who deploys the concept of the past merely in a script-relative way does not.

We can give the primitive tense theorist a name. The idea that temporal structure is articulated fundamentally in terms of the tenses, rather than relations between particulars, finds its clearest expression in the work of Arthur Prior. At the heart of Prior's approach to time is the idea that tense is deeply analogous, both formally and substantively, to modality.

The basic elements of temporal thought or discourse are propositions which change in truth-value over time, and do not need to specify a particular time in order to receive a complete evaluation as true or false. For instance, 'The nectarine is ripe' expresses a complete proposition, that goes from false to true as the nectarine ripens. Temporal information is expressed by means of sentence-level operators that shift the evaluation of a tensed proposition to another time, in something like the way that modal operators

shift the evaluation of a sentence. Prior explains:

If [an] expression constructs a sentence out of one other sentence it is an adverb or an adverbial phrase, like ‘not’ or ‘It is not the case that’ or ‘allegedly’ or ‘It is alleged that’, or ‘possibly’ or ‘It is possible that’...I want to suggest that putting a verb into past or future tense is exactly the same sort of thing as adding an adverb to the sentence. ‘I *was* having my breakfast’ is related to ‘I am having my breakfast’ in exactly the same way as ‘I am *allegedly* having my breakfast’ is related to it, and it is only an historical accident that we generally form the past tense by modifying the present tense, e.g. by changing ‘am’ to ‘was’ rather than by tacking on an adverb. (2003, pp. 12–13)

Formally, the basic elements of tense logic are:

- A stock of propositional variables p, q, r, \dots
- The truth-functional connectives $\sim, \supset, \&, \vee$.
- The four tense operators P, F, H, G .¹⁴

Prior provides no official semantics for his tense logic. A theme that will be taken up below is just how its expressive ideology is related to that of a metalanguage that explicates the meaning of the tense operators in terms of a structure of times related by the earlier-later relation. But the intuitive meanings of the tense operators are: P means ‘it was (once) the case that’, i.e. Pp is true just in case the embedded proposition p was true at some time prior to the time of evaluation; H , means ‘it was always the case that’, i.e. Hp is true just in case p was true at *every* time prior to the present; F means ‘it will (at some time) be the case that’; and G means ‘it will always be the case that’. As with modal logic, the pairs of operators can be defined as duals, i.e. $Pp =_{\text{df}} \sim H\sim p$, and *mutatis mutandis* for F and G .

Prior intended his formal language both as a metaphysically perspicuous representation of temporal reality, and as an accurate formal regimentation of the structures underlying temporal discourse in natural language. The first of these makes it an apt expression of metaphysical views about the reality of tense. One way of understanding the metaphysical territory staked out by Prior’s use of tense logic is the idea to say that

¹⁴In what follows, I consider just the propositional language of tense logic, and completely ignore any of the complications arising from combining tense logic with the logic of quantification.

time is fundamentally unlike space because time, unlike space, is not a dimension of particularity. Rather, tense, like modality expresses a way, or mode, for something to be the case. Apparent discourse about particular events should properly be construed as a product of abstraction over tensed discourse about the subjects of those events. For instance, ‘The death of Queen Anne happened 300 years ago’ is paraphrased as ‘300 years ago, it was the case that Queen Anne dies’. Prior is explicit about this commitment:

What I am suggesting is that what looks like talk about events is really at bottom talk about things, and that what looks like talk about changes in events is really just slightly more complicated talk about changes in things. This applies *both* to the changes that we say occur in events when they are going on, like the change in speed of a movement (‘movement’ is a *façon de parler*; there is just the moving car, which moves more quickly than it di), *and* the changes that we say occur in events when they are not going on any longer, or not yet, e.g. my birth’s receding into the past (‘birth’ is a *façon de parler*—there’s just me being born, and then getting older.) (Ibid., 16–17)

On the second of Prior’s aspirations, it is generally thought that Prior was wrong, and that the expression of tense in natural language is not intensional in the way he thought.¹⁵ Nevertheless, even if natural language does not in fact have the underlying structure of tense logic, but instead quantifies explicitly over times or events, someone might still insist that language is answerable to an underlying level of thought on which this structure disappears, and which has the structure of a tense logic. This would be to see Prior as engaging in ‘logical analysis’ in the style of Russell, rather than empirical semantics; such an analysis might be supposed to dovetail with a correct metaphysics of time. Indeed one of the principle motivations for the A-theory of time is the idea that it fits, at some level, with the basic shape of temporal thought.

In stressing these aspects of Prior’s tense logic, I am departing from one standard way of understanding the formal apparatus. This is the idea that the meaning of the tense-logical formulas is naturally and perspicuously explicated by a metalanguage the quantifies over times. Many attempts to treat tense in natural language using tense logic take this stance, and moreover typically assume that the domain of times forms a fully linear order (e.g. Montague 1973).

¹⁵See e.g. Ogiwara 2007 for an overview of the linguistic evidence that favours an extensional treatment of tense.

It is worth stressing these points because in understanding the tense-logical vocabulary there always remains a temptation to fall back on a conception of time as a linearly ordered dimension, in which particular occurrences are distinguished by their relative position.

Moreover, discussions of the metaphysics of time often assume that the A-series represented by tense logic is effectively just the result of taking a B-series of times ordered by earlier-later and designating a single time as present. This attitude is expressed, for instance, by D. H. Mellor:

The A and B series differ *only* because tenses change with time and dates do not, a difference that reduces to the way B series moments become successively present. Given the inexorable movement of the present along the B series, either series could be defined in terms of the other. (Mellor 1981, p. 23)

Statements of this kind are common, but they do not do justice to the radical nature of Prior's conception of time. Even if it is possible—as Prior argued it was—to capture the structure of time as a linear order by tense-logical means, this cannot be assumed from the outset in the interpretation of the basic notions. The essence of Prior's picture is, rather, that past and future are primitively different ways for something to be the case. And this alone does not imply that past and future generate an order.

A related idea, to be found at various points in Prior's own writings, is that tense logic can be used to theorise about the topology of time, given established correspondences between axiomatic bases of the tense logic and the ordering properties of the earlier-later relation (e.g. Prior 1967, ch. 2, Galton 1984; Newton-Smith 1980). But I shall argue that, despite some convergence, questions about which tense-logical axioms to accept are subtly different from questions about the topology of time.

The crucial feature of tense logic is that its expressions are all *general with respect to time*. An atomic tense-logical proposition varies its truth-value over time; hence, it expresses something repeatable. If we help ourselves to a theory of times in explaining the meaning of the tense-logical expressions, we can express this by saying that a tense-logical proposition functions exactly like a predicate of times. Sticking within the object language of tense logic, we can say that when we have the same propositional letter occurring at multiple levels of scope in a complex formula, for example the letter

p in the formula:

$$(1) p \& P(\sim p \& Pp)$$

there is no good answer to the question whether it is ‘the same’ temporal occurrence expressed by the first and third typographical occurrences of the letter p . The formula says that we have p -ness separated by non- p -ness; but the idea of particular episodes of p -ness, such that we may enquire whether they are numerically distinct or identical, finds no expression in the language. There is no distinction between qualitative and numerical identity provided for here. I will call this feature of tense logic its *generalism*.

This feature has obvious parallels in Russell’s discussion of the two white patches. The point there was that, in describing a situation as one in which white is to the left of white (separated by non-white), we have not yet given any sense to the question whether the two whites are identical or distinct. Doing so requires a conception of them as particulars, and a related distinction between qualitative and numerical identity. With such a distinction in play, one can wonder whether the space of the white patches has an ordinary or a non-standard topology, i.e. whether ‘to the left of’ is irreflexive. But the description of the scene purely in terms of arrangements of qualities does not on its own provide for this.

It is tempting to think that the generalism is cancelled out by the fact that the tense operators encode temporal information in relation to the present. That is, although the propositions of tense logic are in themselves general, they are used in a way that expresses information relative to the time of utterance or evaluation, and thus instances of the same tense-logical proposition can be distinguished relative to that time. In this respect tense logic differs from a fully ‘universalist’ language of tense-logical propositions and a two-place sentential operator BEFORE(____,____).

For instance, formula (1) above distinguishes the two instances of p not directly but rather contextually; the occurrence at different levels of scope in the formula signals their different ‘distance’ from the present, and hence distinguishes the actual occurrences they pick out. (This kind of expression can be facilitated further by introducing metrical operators, e.g. ‘It was the case n time-units ago that...’ etc.) Thus it might be thought that the expression of temporal particularity is to be located not in any formula of the logic, but rather in the way the logic is used, being always centred on the present.

This strategy is much more limited than it first appears. The formula (1) does not in fact necessarily individuate the two occurrences as distinct. Rather, it says that p is in the past of p . We might restate this as saying that p is preceded by p (with an intervening period of non- p .) For this to amount to a judgment of distinctness, we need to assume that nothing is in its own past, or precedes itself. But to do this we already need the concept of a particular, such that the relation of *precedes* can be said to be irreflexive. But it is precisely this notion that the logic lacks.¹⁶ Similarly, the formula

$$(2) p \& Pp \& Fp$$

might be thought to distinguish two instances of p as respectively past and future. But it does not do so: rather, it says that p is both past and future.

This is a structural limitation of tense logic. It looked as though tense logic could supply individuating information not available to the universalist by the fact that, given some sequence of similar occurrences, a tense logical description singles some single one out as being present; everything else is then individuated by their different relations to the present. But in fact tense logic does not do this; although whatever is asserted without being embedded under tense operator is thereby asserted to be present, it is not the case that whatever is asserted within the scope of P or F is thereby asserted *not* to be present. This means, for instance, that the schema:

$$(3) p \supset (Pp \& Fp)$$

is neutral between expressing global recurrence in linear time, and circular time. Indeed tense logic cannot distinguish between these possibilities. Of course failure to

¹⁶In a footnote their discussion of the limits of relationalism about space and time, Hawthorne and Sider attribute richer representational resources to the tense logician in a similar way to that described above: they write “Interestingly, presentists avoid the difficulties considered here: the notion of the present, together with the metrical tense operators (e.g. ‘it was the case 20 minutes ago that’), let the presentist in effect speak of properties had at particular instants of time. Our target relationalist is not a presentist.” (p. 46, n. 26) But this of course assumes that what the presentist’s tense operators are marking out is a linear order with a uniform metric, such that, say, something which happened 20 minutes ago cannot also happen in 20 minutes’ time. More generally, tense-logical statements about distance from the present only have implications for identity on the assumption that what those statements express are relations—whether ordinal or metrical—between particulars. And it is not clear how even to state this assumption in purely tense-logical terms.

distinguish recurrence and circular time is not necessarily a severe limitation, since it is a fairly abstruse question whether these are genuinely distinct global scenarios. The point is just that in this respect it is on a par with the universalist language; contrary to appearances, its present-centredness does not endow it with any extra individuating power.

These points underline the respect in which tense logic makes time analogous to modality. There is a kind of inherent generality in modal thought. A contingent proposition is something that can be true, in principle, at any arbitrary world. This is not to deny that some propositions are true at exactly one world. Indeed each possible world can be associated, or identified, with a proposition that uniquely characterises that world.

Similarly, a generalist tense logic can introduce times, as analogues of possible worlds, by identifying or associating each time with a proposition—intuitively, the conjunction of everything true at that time. Quantifying into sentence position, an instant-proposition i can be defined as any proposition which is both *temporally possible* and *maximal*:

Poss: $\Diamond i$ ¹⁷

Max: $\Box \forall p (\Box (i \supset p) \vee \Box (i \supset \sim p))$

The individuality of each time, on this method, consists just in its being a unique, maximal combination of contingent propositions, which are themselves general with respect to worlds, and recombinable. Since p ranges over all, not just atomic propositions, this allows for the possibility of two times where all the same present-tensed truths hold being nonetheless distinguished, as long as different tensed truths characterise them. Intuitively, this corresponds to the idea of qualitatively indiscernible times being distinguished by their different relations to other times. Of course this method is unable to distinguish particular times if the universe is globally and infinitely repet-

¹⁷Using \Box and \Diamond here for temporal necessity and possibility: intuitively, ‘it is always the case that _____’ and ‘it is sometimes the case that _____’. As long as time is linear, these can be defined: $\Box p =_{df} p \& Hp \& Gp$ and $\Diamond p =_{df} p \vee Pp \vee Fp$. If branching time is possible, then it may not be the case that every time is in the past or future of every other time, and so these equivalences fail.

itive. But one might wonder if this possibility is really distinct from the hypothesis of circular time.

But even given the viability of the tense-logical reduction of times to propositions, it should be stressed that this is a global, holistic method of individuation. Times are distinguished from one another as unique maximal combinations of present states of affairs. If some local repetition is possible, then similar times can only be distinguish with reference to the total global pattern. This plausibly is the case for our intuitive thinking about modality. We normally think of two possibilities as distinct only if there is some non-modal difference between them; thus, the only use we have for the notion of a possible world is that of a maximal possibility, distinguished from other worlds by the totality of what is the case at that world. Moreover, we intuitively distinguish global possibilities just in terms of the *non-modal* propositions that hold at them. It is hard to make sense of the idea of two distinct possible worlds which are precisely the same with respect to their non-modal features, but differ in terms of which modal propositions hold at them.

By contrast, we frequently distinguish between temporal occurrences of the same kind, not by means of contextual information about what else was going on at the time, but on the basis of finding some temporal, or causal, relation between them. There is no obvious analogue of this in ordinary modal reasoning. Our modal thinking does not involve the idea of relations that, in Russell's phrase, imply the diversity of their terms.

The fundamental point here is that ordinary modal thought makes very little use of considerations of order. Insofar any ordinal structure enters our thinking about modality—for instance, in considering the relative similarity of different possible worlds—the relevant relations are grounded in the intrinsic, non-modal character of the worlds. Cross-modal relations do not play any foundational role in our ability to reason about different possibilities. By contrast, our thinking about particular past events obviously relies on being able to discern cross-temporal relations between them. In particular, I have argued that possession of a global temporal framework is connected with an appreciation of causal structure as relating temporally remote events. One might doubt, then, whether Prior's conception of time is really able to capture these ordinal relational aspects of ordinary temporal thought.

So far, in stressing the ways in which tense logic falls short of generating an order in which particulars can be distinguished, I have considered only individual formulas in isolation. But the tense logician has richer resources to appeal to. A common idea is that tense-logical formulas, or more properly schemata, can be used to formulate hypotheses about the topology of time. As emerged in the previous section, there are close connections between questions about the order and topology of a system and questions about the identities of its individuals: for instance, if temporal order is irreflexive and transitive, then going forwards or backwards in time will not return us to the same individual. So, parallel to Russell's suggestion that visual perception can give us the materials to construct a geometry of perceptible space, and thereby distinguish instances of presented sensible qualities, we might look to tense logic to support judgments of the identity and distinctness of temporal individuals by providing the materials for a topology of time.

In the discussion of Russell I argued that this is the wrong order of explanation; in order to think of spatial relations as generating an order, we must already be conceiving of them as relations between particulars. Similarly, I shall argue that, in the present context, interpreting the tense-logical axioms as postulates about the topology of time is subtly question-begging: interpreting them this way is only legitimate given the prior assumption that the tense operators express relations of priority and posteriority between temporal particulars.

6.4 TENSE LOGIC AND THE STRUCTURE OF TIME

As I have emphasised, order plays only a very peripheral role in intuitive reasoning about modality. Nevertheless, as is well known, the axioms of modal logic can be thought of as corresponding to conditions on a relation of relative possibility. In possible worlds model theory, each modal axiom characterises a class of models, specifiable in terms of an ordering property of an accessibility relation on possible worlds, for which the relevant set of axioms is sound and complete. This encourages the idea that, just as the axioms of modal logic characterise conditions on the relation of relative possibility, the axioms of tense logic can be thought of as expressing postulates about the topology of

time, by characterising ordering properties of the earlier-later relation.

The natural way to carry standard modal logic over to tense logic is to translate the modal axioms into corresponding axioms for both past and future. We can give semantic clauses for the tense operators analogous to the modal operators, replacing the accessibility relation R with a temporal relation \leq and its converse, intuitively interpreted as ‘is earlier than or simultaneous with’. Informally:

H: Hp is true at t iff, for all $t' \leq t$, p is true at t'

G: Gp is true at t iff, for all $t \leq t'$, p is true at t'

Assuming, standardly, that P and F may be defined as duals of H and G .

The past-tense analogues of the most common modal axioms are then:

K : $H(p \supset q) \supset (Hp \rightarrow Hq)$

T : $Hp \supset p$

B : $p \supset HPP$

S4 : $Hp \supset HHp$

S5 : $PHp \supset Hp$

(We can save space by assuming a ‘Mirror-Image Rule’ according to which any formula obtained by uniformly replacing all occurrences of P with F and all occurrences of G with H in any theorem is itself a theorem.) We will also need the temporal analogues of the inference rule of Necessitation:

H-Gen: If p is a theorem, Hp is a theorem.

and the corresponding rule for G .

As is well known, the K-axiom characterises the class of all models, T characterises those where R is reflexive, B corresponds to symmetry, S4 to transitivity, and S5 to equivalence. Which of these are appropriate for the natural interpretation of P and H ?

Since K corresponds to no condition on any model, we should expect it to hold whatever the structure of time. We have no reason to question this here. Given that the temporal accessibility relation \leq is interpreted as weak, rather than strict, priority, it is reflexive, and so we should have T . On the other hand B is not appropriate for the relation of earlier-then, except in the unusual case in which time is cyclical. The same goes for S_5 , since it implies B .

Interestingly, there is no axiom corresponding to the denial of cyclical time; that is, the system characterised by the class of models where R is transitive and antisymmetric is just S_4 . Informally, the reason is that there is no expressible difference between a circular model and an infinitely repetitive one. Tense logic does not distinguish between cyclical time and eternal recurrence. More generally, antisymmetry is one of a family of properties that do not correspond to any axiom, including also irreflexivity and intransitivity.¹⁸ One might not be too concerned by this, though. Although there is no standard tense-logical system that rules out cyclical models, there are plenty of formulas that imply time is non-cyclical: any formula of the form $Pp \& \sim Fp$, for instance.

S_4 is closely linked with the principle of the fixity of the past. We might consider the fixity of the past to be expressed by the following schema:

$$\mathbf{F}: Pp \supset GPp$$

That is, the principle that, if something has happened, it will always have happened. This is implied straightforwardly by S_4 combined with the following ‘mixing’ axiom:

$$\mathbf{M}: p \supset GPp$$

S_4 is equivalent to $PPp \supset Pp$, which with M implies F by propositional logic.

Given the clauses for G and H above, and the assumption that P is the dual of H , it follows by quantification theory that M is valid in all models. M says that if p is true at t , all times later than t are such that p is true at some earlier time. As long as that earlier and later are converses, this is obviously valid. Thus, M holds in all models. In other words, the principle of the fixity of the past follows just from the assumption that P and F express a transitive relation of temporal order.

¹⁸For discussion of this point and relevant proofs see Cresswell and Hughes 1984, pp. 47–51.

The main further question about the topology of time that falls within the remit of tense logic is whether earlier-later is a fully linear, or partial, order; that is whether it satisfies the law of Trichotomy that for any pair of times t and t' , either $t < t'$, or $t' < t$, or $t = t$. As it turns out, linearity is captured by the formula:

$$\mathbf{S4.3} (Fp \& Fq) \rightarrow (F(p \& Fq) \vee F(q \& Fp) \vee F(p \& q))^{19}$$

However, as noted above, no formulas correspond to irreflexivity or asymmetry. This has the consequence that we cannot rule out circular time, or distinguish it from eternal recurrence.²⁰ But again, one might not be too bothered by this: perhaps eternal recurrence is a genuine possibility, and it is unclear whether it is really a distinct hypothesis from circular time.

The point I want to emphasise is that there is only a rather limited extent to which choices about the tense-logical axioms correspond to postulates about the topology of time. In particular, as long as we assume that the tense operators do express temporal order, it would seem that the fixity principle F and its mirror-image are not optional. There is no feature of the topology of time that could correspond to its denial. M , as noted, is valid in all models by the definition of the operators. Meanwhile, although $S4$ is valid only in transitive models, it is very hard to see how to reject transitivity whilst holding onto the idea that the tense operators express order.

The question for the primitive tense theorist, then, is what basis we have for accepting M and $S4$, if we do not antecedently take it for granted that tense expresses temporal order. If we really take seriously the idea that tense is analogous to modality, this is a pressing question. For what insisting on an austere, quasi-modal interpretation of the tenses brings out is that there is no *a priori* guarantee that the ‘temporal accessibility’ relation expressed by the operators is a genuine relation of earlier-later. Thus, there is no guarantee that the two relations should be converses, or that it should be transitive.

This point can be further underlined by considering some non-standard interpretations of the operators. In his discussion of the topology of time, Prior acknowledges

¹⁹If we use F^* rather than F , we can omit the last disjunct of the consequent.

²⁰It also means that we cannot rule out models with ‘clusters’: groups of times within the model which are all related in both directions to each other, but only in one direction to times outside the cluster. Cf. Cresswell and Hughes 1984, pp. 81–86.

that ‘non-transitive temporal succession is difficult to imagine’, but suggests the following interpretation:

Suppose time is circular, but as it were changes its sign half-way round. In a cycle taking 3 ‘aeons’, say, to complete, perhaps it will be the case one aeon hence that it will be the case one aeon later that p , but that will bring us to a point which is not itself later but one ‘aeon’ *earlier* than now, so that it is not any interval ‘hence’, but one aeon *ago*; i.e. although we have FFp here, we do not have Fp but Pp . And *only* $A_3 (CFFpFp)$ [i.e., S_4] would fail in such a time-scheme. (1967, p. 39)

What Prior is suggesting here is effectively an reinterpretation of the tense operators where they do not express earlier-later, but a different notion. (In this case, a notion which has both an ordinal and a metrical component.) But once we have admitted this as a possibility, there is no longer any reason to restrict its application to circular models. In *any* ordered structure of times, we can introduce a further relation of ‘temporal accessibility’ over the model, and then use this relation, rather than earlier-later itself, in our semantic clauses for the tense operators. We can choose either to define P and H in terms of accessible earlier times, and F and G in terms of accessible later times; or alternatively we could have two, primitive relations of temporal accessibility, which may or may not be defined as converses. Thus once we distinguish temporal possibility from the order of earlier and later, a number of new possibilities and decision points emerge.

This interpretation of the tense operators, in terms of an intransitive relation of temporal accessibility of limited ‘length’ defined over an order of times, is a nice model for the kind of phase-based thinking described by Campbell. This agent’s conception of time is not strictly circular; rather, in locating itself within a repeatable cycle, the agent ‘sees’ no further back than the onset of that instance of the cycle. The agent’s present cycle is not represented as *identical* to earlier cycles; it just fails to represent any temporal information at all about cycle-instances beyond the present one. Again, it is important to stress that this kind of agent does not represent time as circular. The point is that the way they understand the notions of past and future fails to generate an order.

Consider a different nonstandard interpretation of the tense operators: a semantics involving a branching structure of alternative world histories. The commonest use of this approach is in using forward-branching structures to give a semantics for the open

future; but there is no reason in principle why it should not be applied to the past. The conception of time as a backward-branching structure does seem to have less intuitive appeal than its mirror image—but it is not obvious why this should be so. If the motivation for adopting this kind of structure, for instance, is the idea that truths about other times should be grounded in the present, and if one acknowledges that information about the past may be lost or destroyed, there seems to be no reason in principle to treat the past and the future differently.

There are a number of ways to implement a branching-time logic, involving various decision points. I shall very briefly sketch just one possible way.

Instead of taking truth to be relative to a time t , we instead take truth to be relative to a pair $\langle t, h \rangle$ of a time and a *history*, such that $t \in h$. A history is any linearly ordered set of times, i.e. ordered such that for each pair of times in the history one of the pair is either earlier or later than, or identical with the other. We then adopt the standard semantic clauses for the tense operators, but restricted to times within the history of evaluation. For instance:

P: Pp is true at $\langle t, h \rangle$ iff, for some $t' \in h$ such that $t' \leq t$, p is true at $\langle t', h \rangle$.

We then introduce a further semantic notion of *determinate* truth, or *Truth*, such that a formula is True at $\langle t, h \rangle$ if it is true at all $\langle t, h' \rangle$ such that $t \in h'$ (the h -index for Truth is vacuous.) Determinate truth, rather than truth relative to a possible history, is what agents are fundamentally concerned with: in the context of linguistic communication, Truth is linked with assertability; and Truth connects with action and practical thought insofar as, in deciding what to do, agents are concerned with how things determinately were or will be, not merely how they may be or have been.

This is the most conservative implementation of the branching time strategy; it leaves all of the logic untouched. As long as we are concerned with truth relative to a time-history pair, we keep classical logic with S4 and M, and hence validate F: each individual history is a linear order, and the logic reflects that (in fact, given that we have required each history to be linear, the logic will be S4.3, not merely S4. Without S4.3 the model structures will be more complicated, as we will have to allow for world-histories that are themselves branching; thus we will not be able to define a history as simply any

linear structure, but rather each model will have to designate a class of ordered subsets of the domain of times as its possible histories.)

This kind of structure is invoked by Dummett's anti-realist about the past Dummett 1978b. Dummett considers two possible anti-realist positions, which he calls T and G respectively. The T-anti-realist position is one which is in general realistic, and specifically has a realistic attitude to the existence of present evidence; but it construes the past solely in terms of compatibility with present evidence. If present evidence underdetermines the past, it will be compatible with multiple possible total past histories, each represented by a path in the tree structure. (The G-anti-realist, by contrast, rejects realism about the past on purely general grounds, and makes no use of the idea of a possible history for which classical logic holds.)

Now, as noted, on this interpretation of tense logic all the same formulas are valid as the standard non-branching interpretation. But there is still clearly a difference in the ways of thinking about time that it represents. This difference is not capturable in terms of which formulas are valid at a time, but rather in the way that truths at different times interact. Psychologically, an agent whose thinking was modelled by this structure would affirm all the same temporal principles as an agent with a fully realistic conception of the past; nevertheless, we still want to say they fail to have a grip on the fixity of the past.

As just noted, the formula $Pp \supset Gp$ is valid on this semantics, hence True at any given time. What will fail to be the case, however, is the following principle: if Pp is True at t , then Pp is True at every t' such that $t < t'$. This is because, as we go to later times, these later times will have more backward branches; and p may fail to be true at one of these later-joining branches. The problem is to see how this can be recognised by the agent. At no time will the agent be prepared to assert, or in some other way display a commitment to, the statement 'One day, *that* [i.e. some past occurrence] will not have happened any more'. From their point of view, time is a fully linear order in which the past remains fixed.

We might try to make the difference out by introducing into the logic a determinacy operator, which converts the Truth of a formula into its truth. That is, Dp is true at $\langle t, h \rangle$ iff p is True at $\langle t, h \rangle$. We can then trace entailments, within the object language, between truth and Truth in ways that reveal cross-temporal discontinuities.

For instance, although $DPp \rightarrow DGPp$ is valid, $DPp \rightarrow GDPp$ is not. In general, prefixing each occurrence of a tense operator within a formula with D will not preserve validity.

The question is how to make sense of the meaning of the determinacy operator from the point of view of the agent. It is not as if the backward branches represent some objective diachronic modality, in the way that forward-branching models are often understood to. The obvious interpretation is the anti-realist one, where the branches represent histories that are ‘possible’ in the sense of compatible with present evidence. We might then understand D as a kind of epistemic operator. The way in which, for instance, $DPp \rightarrow GDPp$ fails, on this interpretation, is if there is now sufficient evidence for the fact that something happened, which evidence may subsequently get destroyed. But these are just possible futures on which p has happened for which there is no longer sufficient evidence for p ’s having happened; there is no possible future on which p has not happened. On the other hand, in the future, once the evidence is destroyed, it will no longer determinately be the case that p has happened, and hence Pp will no longer be True for the agent; they will no longer be in a position to affirm, or act on, Pp . In the backward-branching model, this is explained as new pasts becoming possible—new branches joining the structure. But how does the agent think of these now-impossible but soon-to-be-possible pasts? They are no part of what, from the agent’s point of view, is a possible way for things to develop; from that point of view there just is no possible future on which p has not occurred.

It is important to be clear about what the branching structure means. Crucially, it is not intended as a model of the structure of the actual temporal order. It is possible to imagine time as having an actual branching structure (as noted earlier, in Minkowski spacetime the relation of timelike separation is a partial order.) But that is not what is being modelled here. The branching structure as a whole does not directly represent the agent’s beliefs about the structure of time. Rather, the agent’s beliefs about the structure of time, insofar as they have such beliefs, are represented by the structural feature of *each* branch. Insofar as the branches differ, the agent’s beliefs about time are indeterminate.

Thus, on this model, the agent’s conception of the actual history of the world is something that evolves and modifies over time as information is created and destroyed. Although we may use a branching structure to explain the dynamics of these shifts in

perspective, the totality of the branching structure is in principle something the agent never themselves commands a clear view of. Rather, the fact that at different points in time, different world-histories are possible reflects a kind of incommensurability of temporal perspectives.

If this is the correct interpretation of the branching structure, it is unclear how the agent is to understand the *D* operator insofar as it occurs within the scope of higher tense operators. From their point of view, what is determinately the case is just what is the case *simpliciter*; the idea of determinacy as capable of shifting the truth-value of tensed statements within higher temporal contexts requires a kind of self-conscious understanding of the way in which temporal possibility varies between temporal perspectives. But what the model is intended to capture is an agent who lacks this kind of understanding. We are considering an agent whose map of the temporal world continually changes in a way they themselves are unable to represent. I think we should say that such an agent fails to have a proper conception of the actual world. But this does not come out if we restrict our attention to the agent's beliefs at a single time. They may at any given time have quite sophisticated representations of the structure of time. The point is that these are not properly integrated over time so as to constitute a take on temporal reality.

It comes so naturally, when we reflect on the structure of time, to think of it as an order with some kind of intrinsic structure, that it is always tempting to import this conception when interpreting tense logic. But if we take the analogy with modality seriously, this is not justified. If we think of the tenses as primitive labels attaching to repeatable occurrences, there is no guarantee that our use of them will amount to a proper grasp of the temporal framework of the actual world. But, I have argued that our ordinary idea of the past as fixed requires this kind of temporal framework. So tense logic is not adequate to our ordinary ways of thinking about the past.

On the other hand, if thinking of something as past involves locating a particular occurrence in time, then this brings with it the required contrast between the actual and counterfactual. I have argued, in the previous chapter, that historical causal thinking involves just this contrast. This further underlines the sense in which historical causal

thinking concerns particular events.

In the spatial case, I suggested that our grasp of the spatial framework is supported in part by a sensitivity to spatial particularity, reflected in the intrinsic structure of spatial experience. The above discussion suggests we should consider recognising a similar dimension of particularity to temporal experience. This perhaps comes out most markedly in considering the phenomenology of episodic recollection: we can recall particular past episodes, not just the fact that a certain kind of thing happened at some time; it is, plausibly, in virtue of this intrinsic particularity that we can integrate memories of different episodes into a single timeline.

Recognising the intrinsic particularity of temporal experience brings with it commitments in the metaphysics of time. Specifically, it commits us to a distinctive picture of time as a dimension of particularity, and to rejecting the suggestion that time is fundamentally analogous to modality. It also raises questions about the nature of events, and how they exist in time. I will take some of these up in the next chapter.

7 | THE METAPHYSICS OF EVENTS

I have been arguing over the last few chapters that a core aspect of our ordinary thinking about time, specifically thinking about the past, consists in the fact that it concerns particular events. Moreover I proposed, in the last chapter, that we understand this as grounded in the intrinsic particularity of temporal experience. But a popular claim in recent philosophy of mind and action is that events only exist once they are over. For instance, John McDowell writes,

Reality comes to contain, as it were, a particular action of the type describable as ‘my crossing the street’ only when it gets to be true to say ‘I have crossed the street.’ (2011, p. 7)

whilst Anton Ford writes,

As long as someone is closing a door...there is, as yet, no event of the door’s closing. (2014, p. 33)

and Eric Marcus holds,

...events are not present over the whole time that they occur. (2012, p. 217)

Jennifer Hornsby, meanwhile, writes,

When a stretch of ongoing activity is over, an event is on the scene. (2013, p. 9)

And Michael Thompson—in a more formal idiom—has it:

“I am doing A” is no more, or less ‘general’ than “I intend to do A” is; the transition to a genuine particular comes with “I did A”. (2008, p. 137)

For the purposes of this paper, I will take all of the above to be articulations of the following claim:

(E): If an event e exists at a time t , then e is over at t .

(I take ‘over’ here to be equivalent, for present purposes, to ‘complete’, ‘finished’, ‘in the past’, and so on.)

E would seem to imply that particular events do not play any role in characterising temporal experience. Rowland Stout, for instance, in a recent introduction to an edited collection, writes, ‘the standard philosophical accounts that treat actions and experiences as events and states lose, or at any rate misread, the *subjective* aspect of these phenomena, something that can only be captured by thinking of these phenomena from inside the course of their happening.’¹ Given the central role in our experience and understanding of the temporal world which I am advocating for events, it would seem I am committed to denying E.

This chapter will argue that E is incoherent. The basic objection is very simple: events, if they are anything at all, must be the kind of things that occur. But if events are essentially things of the past, this is impossible to understand; if an event only existed once it was over, it cannot ever have occurred. This complaint is flat-footed, yet it cannot be satisfactorily answered. In what follows I will develop this line of thought, and draw out the consequences it has for commitment to an ontology of events.

7.1 THINKING ABOUT EVENTS

It is not clear exactly what is at stake in making a claim like E. The application of ‘exists’ to events is *prima facie* somewhat artificial. Events certainly *occur*; but do they exist?

¹Consequently, philosophers who endorse E often also propose enriching our temporal ontology with an additional category, that of *process*. There is not much unity in the way the term ‘process’ is used by its advocates. Some contrasting recent pleas for processes include Crowther 2011; Galton and Mizoguchi 2009; Hornsby 2012, 2013; Steward 2012, 2013; Stout 1997, 2003, 2016. These different proposals are further discussed and developed further in the papers collected in Stout 2018. Officially, I am happy to be largely neutral about the prospects for processes. However, insofar as theorists of process are motivated by the idea that particular unrepeatable events can play no part in temporal experience, this is something I am committed to opposing.

And if so, does their existence admit of significant tensing? We should not assume that we have a firm pre-theoretical grip on the notions in play here.²

The relevant point for the present discussion is that the nonexistence of events in the present may be thought to impose serious obstacles to the possibility thinking about a particular event while it is going on. John McDowell appears to reason this way when he writes,

By the time there is a particular action of crossing the street for a thought of crossing the street to be directed at, one has crossed the street, and it is too late for the thought to have the nature of an intention. (Op. cit., pp. 7–8)

It is of course not obvious that one cannot think about something that will exist but does not yet exist—if, for instance, one is in possession of a uniquely identifying definite description. Or perhaps there are other, non-descriptive ways in which one might be cognitively related to future particulars—for instance, the relation a sculptor stands in to the statue she will eventually produce. Setting these possibilities aside, our cognitive contact with particulars is typically mediated by our present contact with them. So there is some reason to think that, in the default case, if a particular does not exist yet than one cannot yet think about it; and one might argue along these lines, with McDowell, against the idea that intention in action relates to particular actions.

Of course it might be thought that impossibility of singling out, hence thinking about, an ongoing event is compelling on its own terms—and the quotations from the beginning of this chapter are really meant just to dramatise this, rather than to articulate a prior and independent metaphysical thesis like E. McDowell himself suggests this line of thought in a subsequent passage:

...at the time-bound perspective at which one might say “I am crossing the street”, the logical form of what one is saying cannot await a determination by what is going to happen...So it must be wrong to suppose that [the state expressed by that statement’s] being an intention in action consists in the presence of a *de re* relation to a particular action. (Op. cit., pp. 8–9)

This argument, by contrast, does not appear to go by way of E, but concerns the possibility of interruption and the logical form of progressive-aspect statements. We might

²Doubts about the cogency of a generic notion of existence, applicable uniformly to continuants and events, are expressed in Fine 2008.

articulate this thought as follows: when an event is in progress, nothing at that time determines when its later boundary will be, since the event could be interrupted.³ If events are individuated by their temporal boundaries, that may be taken to imply that nothing determines which of many possible particular events is in progress, hence no individual event can be singled out as the one that is occurring.

But if *E* really is false, such arguments may be less compelling than they appear. When we perceive ordinary objects, not all of their boundaries are present to us at any one time, either. But we do not think this prevents us from perceptually experiencing whole objects; nor does the possibility of perceptual illusion, or cognitive error about an object's actual boundaries, prevent us from singling out bounded objects in normal, non-misleading cases.⁴ At any rate, the status of arguments from perceptual error is highly contentious—and, to make an *ad hominem* point, decisively rejected by McDowell. Similarly, if an event is actually *there* while it is occurring, it is not clear how concerns about the possibility of interruption are much different from familiar attempts to restrict the objects or contents of perception on the basis of considerations from illusion or error about an object's boundaries.

On the other hand, there is an idea in McDowell's discussion which is more broadly consonant with the picture I have been advocating. This is the idea that one's knowledge of what one is doing—knowledge that one might express by saying, for instance, "I am crossing the street" is not yet knowledge of a particular event. On the picture I am developing, this is because knowledge of one's particular actions involves the kind of causal self-location that one can engage in only in relation to one's past actions. But, in contrast to McDowell, I am not explaining this limitation in terms of the idea that, when an event is going on, there is just not enough stuff around yet for one to be able to single it out. I want to allow that, for events other than one's own actions, it may well be that one can single out the whole event on the basis of one's direct perceptual awareness

³Note that this is somewhat weaker than McDowell's way of putting it. McDowell seems to suggest that, if I am crossing the street, there may yet be *no* event for my thought to latch onto. This seems to assume that whatever event does in fact transpire—my crossing of the first half of the street, perhaps—is not something that could have been a complete crossing of the street. Evaluating this assumption would take us into some rather scholastic questions of the cross-modal identities of events, which I do not propose to take up.

⁴A justly celebrated discussion of these issues is Clarke 1965.

of one of its earlier parts.⁵ Rather, on my view it is one's role as agent with respect to the occurrence in question that prevents one from taking the kind of reflective attitude one takes to one's own actions when one locates them in a wider causal framework. And we might understand this in terms of the idea that one's knowledge of what one is doing is not yet knowledge of a particular.

With these points in mind, let me see if there is any cogent and plausible way of interpreting E that would support the idea that events cannot be thought about until they are over. I will argue that there is not.

7.2 EXISTENCE AND QUANTIFICATION

An influential way of understanding claims of existence, due primarily to W. V. O. Quine, is in terms of the first-order existential quantifier. Following Russell (1940), Quine argued that incoherence results when 'exists' is construed as a predicate of individuals, ascribing some property that an entity may possess or lack.⁶ Ontological commitment is revealed instead in the logic of generality: "this is, essentially, the *only* way we can involve ourselves in ontological commitments: by our use of bound variables...to be is, purely and simply, to be the value of a variable." (1948, pp. 31–32) On this conception, the quantificational form 'there are *F*s' has priority over the predicational form '*a* exists'.

The Quinean conception of existence would seem to secure the connotation that only existent entities can be thought about. Setting aside complications concerning fictional characters and so on, it is plausibly a mark of singular thoughts and experiences—that is, thoughts and experiences which relate subjects to particulars—that they can appropriately be described in such a way such that reference to the object of thought or

⁵There may yet be a disanalogy with the perception of whole objects. It is not clear that, when one sees a whole object, one does so on the basis of some *more direct* relation of awareness one bears to its near surfaces. Rather, one might instead take the view that perceptual awareness generally distributes over part-whole, in much the same way that spatial locatedness does, with no particular explanatory priority. (For further discussion of these issues, see Martin 2017.) On the other hand, one might think that the temporal character of perceptual experience is such that all one immediately *sees* is that portion of an event which occupies one's experienced present; but that one indirectly sees the longer event of which it is part, or otherwise can cognitively single it out in virtue of one's perceptual awareness of a part of it.

⁶See, however, Moore (1936), who makes the case that there is no difficulty in supposing that 'exists' is simply a predicate that applies to everything.

experience is open to existential generalisation. For instance, if I am thinking about a particular book, it follows that there is a particular book about which I am thinking. To put this another way: in order for a subject to enjoy thoughts and experiences of a particular, there must be such an item as the object of their thought and experience. If there were no such item, then there would be nothing for them to be thinking about or experiencing.

The Quinean conception of existence dovetails nicely with one influential source of motivation for an ontology of events, namely the semantic structure of sentences that report actions and other goings-on. Davidson argued that these sentences are to be analysed as involving an existential quantifier ranging over events.⁷

A key observation of Davidson's is that a sentence such as,

- (1) Guthrie fell over.

asserts that Guthrie fell over at least once, but not how many times he fell over. However, it is subject to adverbial count-modification, providing this lacking information, e.g.:

- (2) Guthrie fell over three times.

These facts, among others, suggest that we should think of sentences like 1 and 2 as a kind of existential quantification—quantification over events. Hence 1 is analysed as:

- (3) $\exists e(\text{Fell over}(\text{Guthrie}, e))$

i.e., that there was an event of Guthrie's falling over; and 2, similarly, as the statement that there were (at least) three events of Guthrie's falling over.

It is notable that this analysis is most readily applicable to past-tensed sentences. Specifically, the main present-tensed counterpart of the verbal phrase *fall over* is the progressive-aspect:

- (4) Guthrie is falling over.⁸

⁷E.g. Davidson 1980c,d. Subsequent developments of Davidson's event-semantic programme include Higginbotham 2009; Parsons 1990; B. Taylor 1985; Wilson 1989.

⁸The simple present 'Guthrie falls over' is normally—although not universally—given a 'habitual' reading, relating to a general pattern or tendency rather than a specific occurrence.

This, however, is not susceptible to the same count-modification:

- (5) Guthrie is falling over three times. (?)⁹

This difference between perfective and imperfective aspect is suggestively analogous to that between count- and mass-quantification. This has led some to suggest that the progressive aspect does not quantify over events, but rather over stuff-like processes.¹⁰ Moreover, the progressive exhibits the so-called ‘imperfective paradox’: statement 4 is not proven false if Guthrie does not in fact fall over (being propped up at the last minute by a helpful companion.) Thus it seems that the truth of 4, unlike that of 1, is independent of the actual existence of any event of Guthrie’s falling over.

These brief considerations suggest that the past tense, but not the present, is appropriately understood as directly quantifying over events; and this may be thought to provide some support for E.¹¹ If we understand existence along Quinean lines, then the thought is simply this: past-tensed perfective verb forms, but not the present tense, involve quantification over events. Since existential commitment is revealed only by quantification, we should thus only recognise the existence of those events which are reported by the past perfective. Given that this notion of existence satisfies both our constraints, it thus seems to follow, as per McDowell’s argument, that there is no thinking about an event until it is over.

But this line of thought is not conclusive. In particular, it is questionable whether E can be fully understood just in terms of the quantificational structure of verbal statements. The following section will develop this point.

7.3 GENERAL AND SINGULAR EXISTENCE

The Quinean analysis applies in the first instance to *general* existence claims. Indeed, the Russell-Quine view is motivated at least in part by a desire to avoid the difficulties

⁹On the other hand, ‘Guthrie is falling over for the third time’ is perfectly acceptable. This suggests that the linguistic data may be more equivocal than they initially appear; but I will not pursue this line of enquiry here.

¹⁰Observation of this analogy is due principally to Mourelatos 1978; cf. also Bach 1986.

¹¹Marcus in particular seems to motivate the claim in this way: for instance he writes, “If no one has crossed Columbus Avenue, there is not yet a crossing of Columbus Avenue, i.e. the answer to the question ‘How many times has Columbus Avenue been crossed?’ is ‘none.’” (op. cit., p. 217)

associated with singular existentials. But it is not clear that, in the present context, questions of singular existence can really be circumvented. Consider again, for instance, Ford's version of **E**:

As long as someone is closing a door...there is, as yet, no event of the door's closing.

Understood strictly and literally, this claim is plainly false—unless, perhaps, the door is being closed for the very first time. What Ford is trying to say is presumably something like: if someone is closing a door, there is not yet such an event as the one that, if they finish closing the door, will then be identifiable as *that particular event* of the door's closing. But this seems to throw us straight back into the puzzles of singular nonexistence that Russell and Quine were at pains to avoid. For what is the event whose existence is being denied, if there is no such event?

Considered just as a difficulty of formulation, this problem might be overcome. Ford could perhaps have said that, if someone is closing a door (or, indeed, if they are not), then the number of events of the door's closing is equal to the answer to the question 'How many times has the door closed?' But the difficulty goes deeper than this. The idea behind **E** is not merely to express the generalisation that, for all the events, each has the property of being over—this would be like the claim that miracles are all in the past—but rather is a thesis on the temporal nature of events, concerning *when they come into existence*: namely, at the point of completion.

We can bring out the difference between these claims by relating **E** to a general principle about existential claims: that, if true, they have true singular witnesses. Thus:

∃-Witness: For any true statement¹² of the form ' $\exists x\varphi(x)$ ', there is at least one true statement of the form ' $\varphi(a)$ ', where ' a ' is schematic for a term denoting some individual.¹³

¹²I follow Rumfitt 2010 in using 'statement' to mean an interpreted sentence-type with all (truth-conditionally relevant) contextual parameters fixed. As will become apparent, my use of the notion is highly idealised, such that there may be a true statement of the form ' $\varphi(a)$ ' even if no actual practice exists of using a name which can replace ' a ' to yield a true sentence. Perhaps this notion of a statement might be glossed as an interpreted *possible* sentence, although there may be some difficulty in precisely spelling out the kind of possibility involved; alternatively, one may prefer to talk about language-independent propositions.

¹³N.B. \exists -Witness fails to hold in some non-standard logics, notably supervaluational logics. A full

As applied to the event-existential analysis of action sentences, this means that a statement such as 1 above requires for its truth a singular witness; a true statement of the form:

- (6) 'e is an event of Guthrie's falling over.'

or, more generally,

- (7) 'e is a φ -event.'

where 'e' is, again, not a variable but a schematic term going proxy for a *name* for an event. We can regard the commitment to there being such singular witnesses as partially definitive of the proposal that action sentences involve existential quantification. This commitment is easy to miss, since genuine names for events in natural languages are few and far between—with notable exceptions for especially momentous events, as in 'Woodstock', 'Watergate', 'The Great Ordovician Biodiversification Event'—and hence statements with the form of 7 are rarely actually asserted. In this respect, an ontology of events digs below the surface of everyday linguistic practice; in construing a significant fragment of language as involving covert quantification, it posits a layer of unasserted singular truth lying beneath ordinary discourse about actions and events.¹⁴

This departure from linguistic appearances is not, in itself, anything the proponent of E should object to. Unless they hold that E is vacuously true, they will allow that very many events do exist in the past, and hence allow that there are vast numbers of

discussion of the prospects for a supervaluationist interpretation of discourse about temporary existents would take me too far afield. However note that my concern is fundamentally with the conception of time as a system of particulars. I am arguing that such a conception is supported by the cognitive contact we enjoy with particular events. It seems clear that a merely supervaluational interpretation of event-existence claims is not adequate to support such a conception; the point I urging is precisely that our contact with events is not just the awareness of the existence of some event or other of the relevant kind, but awareness of a particular event. We ought, for instance, to be able to introduce names for such events, if we so desired.

¹⁴Of course Davidson would take this to be at best a misleading description of the commitments of his event semantics, since he upholds Quine's doctrine of the inscrutability of reference, and hence takes the notion of a singular statement, articulated into name and predicate, to be ultimately explicable in terms of the holding of entailment relations between sentences of a language with quantificational structure (Cf. Davidson 1984b,c.) In this respect I am departing from Davidson's Quinean heritage in asking after the singular bases of existential truths; but this, as I argue, should be common ground between me and my opponent. I take this point up again in section 6.

entities for which we typically lack individual names but are nonetheless committed to. Indeed this may not be such a departure from appearances in a wider sense, if, as I am arguing, particular events have an important role to play in our cognitive economy—for instance, as the objects of episodic memory—albeit one that is not typically expressed linguistically through the use of proper names.

Given their acceptance of particular events, then, the proponent of *E* claims that, for each individual such event, it came into existence only at its point of completion; before it was over, it did not exist. This is crucially different from the claim that all presently existing events are now over; it is different even from the claim that, always, all events are over. Rather, it is at least as strong as the claim that, always, all events are *always* over.¹⁵ In other words, it is distinctive in denying, *de re*, the existence at one time of those very events which, at a later time, are recognised to have occurred. This cross-temporal identification of individuals constitutes a level of structure that emerges only once we get into view the singular truths about events that witness the general ones. For any true past-tensed statement such as 1, we consider any particular event such that its being the event that it is—its being an event of Guthrie’s falling over, say—suffices for the truth of that statement. We then consider what to say about this event at the time before it was complete. The claim of *E* is that, at such a time, this event did not exist.

This understanding of *E* is required once we realise that existential claims must have singular witnesses. But it is also required by the dialectical context in which it is made—for instance, in the argument against *de re* intentions here attributed to McDowell. The alleged problem for singular thought about ongoing actions is not that, while the action is ongoing, there are no events, or not enough events, of the relevant kind—the former is generally speaking false, and why should the latter be relevant?—rather, it is that the very event which is the sole candidate target of a *de re* intention, the one which is currently underway, is not yet complete and hence does not exist. Such arguments thereby require us to trace events back to that liminal phase of their careers before, as *E* has it, they emerged into reality as full-blown particulars.¹⁶

¹⁵The latter two claims may diverge in truth if we allow, as the defender of *E* presumably should, that the domain of existing events may vary over time. I say ‘at least as strong’ because one might take *E* to be a claim about the essential nature of events that is not reducible to, though implies, existence claims expressible in quantified tense logic.

¹⁶McDowell expresses a recognition that there may be a difficulty here when he writes, “things...that

Hence **E** is irredeemably mixed up in the shady business of singular nonexistence claims.

This is not yet an objection. The disreputability of singular nonexistence claims may not be deserved; indeed the following section will suggest a way of making sense of them, at least in temporal contexts. The present problem is just that the Quinean conception of existence is inconsistent with what the defender of **E** needs to say.

Singular (non)-existence claims require us to understand ‘exists’ as applying to individuals; according to Quine, it is a quantifier. We can, however, generate a predicate expressing Quinean existence by using the identity predicate: the open sentence

$$(8) \exists y(x = y)$$

says that x is identical with something; that is, there is such a thing as x . Given classical logic with identity, the resultant property is one which provably applies to everything.¹⁷ So how can there ever be anything to which it fails, or has ever failed, to apply?

This notion of singular existence—existing as simply being something—is what Timothy Williamson calls ‘existence in the logical sense’. As he puts it, ‘Nonexistence in the logical sense is a very radical matter indeed, for it entails having no properties or relations whatsoever.’ (2002, p. 246) As such, it is hard to understand how something could fail to possess existence in the logical sense. This threatens to render incoherent the singular nonexistence claims required by **E**.

More precisely, if the logic of our quantifiers is classical, with unrestricted Existential Generalisation and standard logic of identity, the Quinean conception makes existence a consequence of anything whatsoever’s being true of an object—since from any statement about any object a , we can infer that there is such an object as a . An immediate consequence of this is that the supposition that a does not exist is contradictory; since from the supposition that there is no such thing as a , it follows that there is some x such that there is no such thing as x .¹⁸ Indeed, given quite weak assumptions

do not exist yet are not yet the particulars that, we may awkwardly say, they are going to be.” (ibid., p. 8) The following few paragraphs show how the awkwardness in McDowell’s chosen mode of expression becomes a full-blown contradiction as long as we conceive existence in Quinean terms.

¹⁷If everything is self-identical, it follows by Existential Generalisation that everything is identical with something.

¹⁸That is: $\sim\exists x(a = x)$ entails, by EG, $\exists x\sim\exists y(x = y)$, which by the logic of identity entails the direct contradiction $\exists x(\exists y(y = x) \& \sim\exists y(y = x))$.

in quantified modal and tense logic, it is provable that everything exists—in the logical sense—eternally and necessarily.¹⁹

Williamson takes such considerations to establish that everything does, indeed, exist necessarily and eternally. If this is right, then of course *E* is false. But one might alternatively conclude that there is something wrong with understanding temporary existence in these terms. At any rate, it seems that the logical sense of existence cannot be what the defender of *E* has in mind.

An obvious response here is to reject classical logic and adopt a free logic, on which there can be significant truths about non-existent entities, with Existential Generalisation restricted accordingly to existent entities. This move has the consequences of, first, making singular nonexistence not actually contradictory, hence blocking any formal proof of eternal and necessary existence; and, secondly, allowing temporal discourse more generally to range over entities at times other than those at which they exist. In these respects a free logic seems like a good candidate for a formal regimentation of discourse about temporary beings.²⁰

The problem with this move is that it threatens the idea that one cannot think about nonexistent entities. If there can be truths about nonexistent entities, what is the significance of their nonexistence—why couldn't thinkers have thoughts, perhaps even experiences, of entities that do not presently exist? On a Quinean conception of existence combined with classical logic, absolutely nothing can be true of a nonexistent entity—not even that it does not exist. There is thus a purely formal route from the idea that I thinking about *x* to the idea that *x* exists, since being thought about by me is a condition that *x* satisfies. But if we relax the logic, this purely formal implication is no longer valid. Thus we cannot argue contrapositively, from the nonexistence of *x* at a certain time, to the impossibility of thinking about *x* at that time.

This challenge raises questions of how the free logic is to be interpreted—in par-

¹⁹This result follows from the notorious Barcan Schema and its converse. BS is provable using only classical logic, Necessitation the B-axiom; the Converse Barcan Schema requires only the weaker K-axiom. From these it follows (again with classical logic) that all possible individuals actually and necessarily exist. The temporal analogue is that anything which exists at any time, exists eternally. For a helpful overview of the interaction between these technical issues and philosophical questions of temporary existence, see Sullivan 2012a.

²⁰For a free logic for temporary existents, see Cocchiarella 1991.

ticular, whether the name position in a true singular negative existential of the form ' $\sim\exists x(a = x)$ ' is understood as denoting a non-existent object. A purely negative free logic can validate such statements simply by assessing all atomic statements containing empty names as false, and then treating truth-functions classically.²¹ The natural interpretation of this is that atomic sentences with empty names are false, and their negations true, because not about any individual at all. A negative free logic of this kind can validate the implication $\sim\exists x(a = x) \Rightarrow \sim\varphi(a)$, as the contrapositive of the principle $\varphi(a) \Rightarrow \exists x(a = x)$, where φ ranges over only *atomic* predicates.

The present problem, by contrast, concerns the intelligibility of saying, *of* some given individual event, that it did not exist at a certain time. This requires us to interpret the negative singular existential, not as truly denying a statement which is not about any individual, but rather as truly withholding the property of existence from a given individual. Minimally, the once-nonexistent event is identified with one that now does exist; this is a true identity, hence a true atomic statement. Moreover, arguments like the one attributed to McDowell require us to consider an event at the time before it was complete, when it was still underway, in order to then deny its availability for thought at that time. This requires us to affirm the positive statements that, for instance, the event commenced at a certain time, and so, at the relevant time, it was in progress. So our free logic must be one which allows true *atomic* predications of nonexistent entities.

Thus the question arises, What is the property we are withholding when we deny that an event exists at a time? And why should lacking it make a difference to what thinkers at that time are in a position to think about? An equivalent way of putting this thought is: adopting a free logic in this context looks a lot like just imposing a temporal restriction on quantification. We simply restrict the range of the quantifiers, at a given time, to entities which enjoy a special ontological status at that time. But if this is all it amounts to, why couldn't we simply *drop* the restriction, and quantify over ongoing and future events?²² What is the special ontological status of those individuals which are held to exist at a given time, on the basis of which they, but not others, are included

²¹See e.g. Burge's (1974) free logic for singular terms.

²²Indeed free logics like Cocchiarella's have in their semantics an 'inner' domain for each world or time, comprising the individuals which exist at that world/time, and an 'outer domain' which is the union of every inner domain. Hence it is straightforward to introduce a second pair of quantifiers, whose range at each world is the outer domain, and whose logic is classical.

in the range of the free logician's quantifiers at that time?

The reply might come that we are not in a position to quantify over ongoing (or future) events; that our forms of temporal representation concern only those particular events which are in the past, and *E* is meant to reflect that fact.²³ But, again, we need some explanation of why this should be the case, if there nonexistence does not amount to their sheer absence from the universe, from the totality of what is the case. It is inadequate at this stage of the dialectic to reply that temporal quantification is normally restricted only to those entities which can be thought about at the time in question. If we interpret *E* in this way, then any argument from it to the impossibility of thinking about ongoing events looks flatly premise-circular.

Absent some further account of what it is for an event to exist, moving to a free logic remains a purely formal manoeuvre, without the substantial connotations required to validate arguments for the unthinkability of ongoing events. We need some independent, substantial notion of event-existence, thicker than whatever classical quantification theory provides, susceptible of significant tensing and bearing systematic explanatory connections with singular thought and experience. In the following section I offer such an account of the existence of continuant substances, before returning to events.

7.4 SUBSTANTIAL EXISTENCE

Many substances are perishable. Living things, for instance, are born and die, and we want to be able to acknowledge that birth and death are a kind of coming-to-be and passing-away. Yet we are also obliged to admit truths about living things at times after their death (Frege was celebrated only after his death) and we quantify over past things that were not simultaneously present (there were 13 emperors of the Qing dynasty.)

What this seems to suggest is that, at least for continuants, we should acknowledge a kind of existence which is stronger than merely being a possible value of a bound variable, one which also admits of tense. Call this existence in the *substantial sense*. This substantial sense of existence has connotations of certain ways of behaving and

²³Marcus writes: "To abandon this principle [that there is only an event of e.g. someone's crossing the street once they have crossed the street] would be to surrender our grasp on the identity and individuation conditions for this event type." (op. cit., p. 217)

concretely being in the world, which may be more or less specific to things of a certain kind. In the case of living things, it is plausible that existence in the substantial sense is intimately connected with *being alive*, and with the kinds of living activity connected with that thing's species.

This may seem to threaten to make 'exists' ambiguous, with two relatively unconnected senses. But they need not be unconnected: for temporal things that enjoy substantial existence, to exist in the logical sense can be identified with existing in the substantial sense at some time or other.²⁴

In the case of continuant substances, it seems plausible to maintain that they are unavailable to thought before they come into substantial existence—before they are born or made—but in principle available thereafter. The conception of existence sketched here affords at least the outline of an explanation of these intuitions, in the following way: thinking about an individual requires, at some point, singling it out. More specifically, thinking about an individual requires either singling it out there and then, or being appropriately causally related to a prior act of singling it out, for instance through memory or through competence with a proper name whose use can be traced back to some initial act of singling out. Now, plausibly, something can only be singled out if it is *there*—if it is actually concretely present there and then for the singling out to take place.²⁵ This is just to say that entities can only be singled out at a time if they exist in the substantial sense at that time. From this it follows, assuming that causation is unidirectional, that as far as temporal continuant substances are concerned, the only thinkable entities are ones that substantially exist at the present time or prior to the present time.²⁶

²⁴A treatment of tensed existence along these lines is given by Woods 1976. Compare also Anscombe 2015, p. 297, Geach 1954, pp. 266–268, Wiggins 2001, p. 69. All the above might be regarded as semantic elaborations of the Aristotelian idea that existence, for substances, is the actuality of a certain kind of *activity* or active being (cf. *Metaphysics* Θ; for recent expositions see Beere 2012; Kosman 2013.)

²⁵This is quite a weak claim. It is much weaker, for instance, than the sortalist thesis that singling out requires having an adequate conception of the singled out thing's kind. However, there is clearly much more that needs to be said about the notion of singling out than I can attempt here.

²⁶This is a rough generalisation. Perhaps there are exceptions: perhaps it is possible to think about an artefact while it is under construction. Or perhaps I can think about my future ham sandwich before the upper piece of bread is in place. But what will still be true is that the possibility of thinking about something before it substantially exists will depend on special considerations about the kinds of processes by which a thing of that kind comes into existence, the ways in which its constituent matter is assembled and transformed, and the ways in which a thinker can be party to those processes. Thus there are still broad temporal constraints on thinkability imposed and explained by the notion of substantial existence.

Thus for continuant substances we have a notion of substantial existence that can play two key theoretical roles. First, it can explain the sense in which substances come to be and pass away, but without making singular nonexistence contradictory. The Quinean quantificational conception of existence can still stand, but we now understand it in terms of substantial existence: in our temporal sophistication, we have a grasp on the idea of things substantially existing at times other than the present, and can extend our (classical) quantifiers to include those past and future entities.²⁷

Secondly, substantial existence has an epistemic and temporal significance, since cognitive access to particulars is enabled by acts of singling out. Access to particulars which are not presently (substantially) existent is hence mediated by causal connections to episodes of contact with them at times at which they were. Because thought about particulars is thereby bound up with the workings of causality, it bears causality's temporal orientation. This introduces a fundamental asymmetry between our access to past and future individuals. The question to now be taken up is whether this style of explanation might apply to events, in a way that would support E in the present context.

7.5 BEING AN EVENT

Building on the previous section's sketch account of the existence of continuants in time, E might be understood as saying: there is some substantial mode of being for events, importantly analogous to the active being of a continuant substance, associated with *having occurred*. This makes events analogous with continuants in the following way: the completion of an event is something analogous to the birth or creation of a continuant; the period leading up to the completion of the event is analogous to the process of gestation, assembly, or whatever, which leads up to the creation of a new continuant. Moreover, for events, *having occurred* places constraints on cognitive access similar to those that, for instance, *living* places for biological continuants.²⁸

An essential aspect of the picture just sketched for continuants is that we only under-

²⁷I leave it open whether an analogous account can be given of the priority of actual over merely possible existence.

²⁸This analogy is especially prominent in Thompson and McDowell. Cf. Thompson, *op. cit.*, pp. 134–137; McDowell, *op. cit.*, pp. 8–9.

stand the mere logical existence of a continuant in terms of its past or future substantial existence. The substantial sense of 'exists' is its *focal* sense, underlying the thinner, logical one. Napoleon's existence as a mere subject of predication is not a matter of a kind of logical shadow obtruding into the present; rather it consists in the fact that he *once* existed as a flesh-and-blood human, although he no longer does.

This is connected with a more general point about the basis of our concept of the past. In thinking about things past, we have a sense of something somehow actual, but nevertheless absent from the world as we now experience it. Central to this conception is our appreciation that the past comprises the same kinds of things, broadly speaking, that we can experience now; and the very same things, strictly speaking, that were once available to be experienced. A full account of how this degree of cognitive sophistication is attained would likely invoke the capacity for episodic recall as a special and canonical mode of access to the past; but these are questions for another paper. The key point is that it is only thanks to our grip on the idea of past presence that we can understand the idea of present pastness. The primacy, in temporal contexts, of the substantial sense of existence, is a consequence of this conception. It is because we can temporally 'decentre', and understand the idea of there being, say, a living human being at a time now past, that we can understand the idea of a merely past human.²⁹

But the idea that *having occurred* could supply a focal sense of existence for events destroys this conception of the past. It requires that we think of the past as comprising entities of a *sui generis* kind, categorially different from anything to be encountered in the present. Not only are these entities unencounterable, but we cannot understand what it would be for such a thing to be present to us: their being consists precisely in their absence. If this is what events are, then we lose our grip on their reality altogether. The domain of past events becomes a mythical realm, removed from any conception of actuality we can form.

The analogy with substances is in this respect profoundly misleading. When a substance comes into being, we can distinguish process and product: the process of coming-to-be, and the new individual that results. The possibility of the distinction rests on the fact that, once a substance is created, it is *there* to be met with, part of the present fab-

²⁹This is essentially another way of putting what, back in chapter 2, I called the dependent character of our thinking about the past.

ric of the world; we thereby separate it from the goings-on, now over, that led up to its existence.

But attempting to resolve non-substantial changes, like my crossing the street, into process and product, is hopeless. What is the thing produced when, say, I cross the street? Just: the event of my having crossed the street. But if this is distinct from the process by which it came to be—from what was actually going on while I was crossing the street—we are left utterly in the dark about what the event, as so distinguished, might be.³⁰

A corollary of this is that it becomes equally obscure what ongoing change in the present is, if not that which will have occurred once present activity has ceased. The introduction of processes, as distinct from events, responds to a felt need for something to make up for the absent event. But once events are banished from the present, invoking such surrogate entities cannot help. Consider again my crossing the street. This event is something that involved, for instance, a continuous change in my position from one side to the other. The question is what it was for this to be happening, for my position to be actually changing in that way. The process cannot be identified with any specific extended change, for instance the transition from my being on one side of the street to my being on the other, for then it would just be an event.³¹ So the process must be something additional to the actual changes that occurred—perhaps from which the changes are constituted, or which causally gives rise to them. It may be that there is some such legitimate notion of process along these lines. But it will not help us understand what present change is if we insist on separating out the change that will have occurred from the process that constitutes it or brings it about. As long as processes are distinct from events, they cannot plug the gap left by the absence of events in the present.³²

³⁰This point is appreciated by O'Shaughnessy: "...the on-going of a process is not a *distinct phenomenon* from the happening of the event it constitutively realises. How can the constituting of *an event*, as opposed to a *material object*, be distinct from the occurrence of the event? The object moves off into other times, and perhaps other places, and in this sense *transcends* its spatio-temporal beginnings, but the event does not...[an event] cannot cut itself adrift in the manner of an object. Failure to grasp this difference can make the division between occasion and content within the event look like a split. In modelling the event upon the material object, we set its content adrift, but it is a drift that leads it *out of time*." (O'Shaughnessy 1971, p. 222) Puzzlingly, O'Shaughnessy nevertheless insists that the process which he calls the 'constituting' of the event is somehow distinct from the event itself.

³¹Cf. Steward 2012, p. 383: "Individual processes, unlike events, are not changes."

³²Marcus responds to this problem in a slightly different way. He says that the actuality of change in

Denying the existence of ongoing events opens up an ontological rift between the past and the present, and thereby renders their connection opaque. If the past comprises its own, proprietary kind of entity, foreign to the present, we can no longer regard the past and the present as of a piece, the past having once been present and the present soon to be past. We thus lose any sense of a single, unified temporal reality, of which past and present are complementary faces. We are set adrift in time, stuck in an unmoored, free-floating present.

7.6 EVENTS DEFLATED?

The previous section pressed the requirement that our understanding of what it is for something to exist in the past must in some way derive from, or be grounded in, our understanding of what it is for such a thing to exist substantially in the present. But the defender of *E* may reject this requirement as spurious. They may reply that the assimilation of events to continuants has been miscarried: although our understanding of the past existence of continuants is grounded in our understanding of their present existence, this is not how it is with events.

How, then, is the idea of an event's past existence to be elucidated? The answer offered may be: our understanding of the existence of events constitutes no more than what is manifested in our competence with past-tensed aspectual verb forms, specifically with the alternation of imperfective 'A was φ -ing' and perfective 'A φ -ed'. For instance, Sebastian Rödl writes,

...a thought refers...to a concept with an argument place for events if and only if it predicates...a movement form. We grasp the contrast of event and event-concept through the contrast of movement and movement form, which contrast is grounded in the contrast of aspect. (2012, p. 164)

Abstracting from Rödl's specific terminology, the key idea in this passage is the suggestion that the category of event, or temporally extended occurrence, is founded in the

the present 'consists not in the existence of a particular but rather only in the obtaining of a fact'; a fact which he goes on to explain as a kind of potential, 'a forecast for new particulars.' (op. cit., p. 218) It is not clear how this helps: the fact of a certain potential for an event, in the present, does not on its own amount to the actuality of present change.

contrast of imperfective and perfective aspect that certain verb-forms admit.³³ Accordingly, one might extract from these remarks the suggestion that we can understand the idea of a particular event of, say, my crossing the street, existing in the past just in terms of the truth of the pair of statements ‘I was crossing the street’ and ‘I crossed the street’.

This response embodies a certain deflationary attitude to the category of event. We understand the past existence of an event in terms of the truth of aspectual past-tensed statements, rather than vice versa. Thus the demand for any further account of what it is for an event to exist in the past is rejected; insofar as existence in the past is intelligibly related to existence in the present, this goes via the relation of the past perfective to the present progressive—where the latter, unlike the former, does not introduce any particular event.

The position staked out by such a response is, I think, at least coherent. The question to be asked is whether it really amounts to an account of what it is for an event to exist, or rather a rejection of events as genuine entities. Davidson’s idea was that an analysis of action sentences in terms of events is somehow a more perspicuous representation of their logical form than the surface appearance of a noun phrase joined to a verb. By contrast, the deflationist’s suggestion is that statements about events be understood ultimately in terms of aspectual verb forms predicated of substances. This recalls A. N. Prior’s insistence that ‘what looks like talk about events is really at bottom talk about things’ (2003, p. 16).

In order properly to evaluate the deflationist’s position on events, we need some surer criterion of being genuinely committed to an ontological category. And such a criterion is provided by the principle of \forall -Witness introduced in section 3. If one accepts that there are true existential statements about past events, one is thereby committed to these statements’ having singular witnesses—to there being singular truths about past events of the form ‘ e was a φ -event’. But the deflationist is not able to offer any paraphrase of these statements in terms of their aspectual verb-forms. All that ‘A φ -ed’ gives us is that there was at least one φ -event; it does not identify any particular φ -event as making such a statement true. This lack of a way of expressing singular event-

³³Cf. also Thompson, pp. 21–22. A formal elaboration of this idea in the setting of a tense logic in the style of A. N. Prior is to be found in Galton 1984. Both Thompson and Rödl draw heavily, albeit critically, on Galton’s ‘event-logic’ in developing their accounts of event-forms.

statements should indicate to us that the deflationist is eliminating, not explaining, the category of event.

In pressing \exists -Witness as a criterion of commitment to events, I am of course rejecting the Quinean idea that ontological commitment is expressed solely through the use of quantifiers. But, as I have argued, the Quinean conception of existence is not one that is conducive to the philosophical ambitions of *E*'s proponents. In particular, *E* is here understood, in the context in which it is made, to theorise events as entities whose temporary existence places systematic constraints on when it is possible to think about them. This exerts pressure towards an account of events, and of how they exist in time, that goes beyond what is expressed by the fragment of language in which they feature only as possible values of variables; we need a more substantial account of what it is for an event to exist that connects up with questions about singular thought and experience. An account of this sort for continuant substances was sketched in section 4. But such an account of the existence of events is precisely what the deflationist declines to provide.

As such, insofar as the deflationary position offers a stable fallback position for the defender of *E*, it is not one that satisfies the need for a coherent conception of temporary existence with systematic explanatory connections to singular thought and experience. This fact brings out sharply the conflicting philosophical impulses of which *E* is an expression. We can see its apparent plausibility as arising from a combination of a Quinean approach to ontological commitment—one that views ontological questions as settled on the level of generality, for instance by questions of quantificational structure—and a distinctly un-Quinean concern with how particular things exist in time, and with how their temporal nature shapes their role as objects of thought and experience.

7.7 CONCLUDING REMARKS

We may perhaps see the motivation for *E* as arising, ultimately, from a misplaced attempt to make sense of our relation as agents to a wider temporal world. I have argued that part of what is distinctive about the past is the way that we are able causally to locate ourselves in the past, where this self-location is supported by the knowledge we enjoy of particular past events. And I have argued that our involvement as agents in what

we are doing prevents us enjoying this same knowledge in relation to our present and future actions. To this extent, I am sympathetic with the general thrust of McDowell's idea that intentions in action do not relate *de re* to the particular actions that they are intentions in. But, I am also insisting that we understand this feature of our perspective as agents in terms of a limitation on how much we can know about our own causal histories, in virtue of our agential partiality. The fundamental mistake involved in a claim like E, I suggest, is to understand this limitation in terms of a categorical difference between past and future entities. Understanding our perspective in this way makes the past into a disconnected and autonomous realm, and thus renders our belief in it ultimately mythical.

The difficulty is to make sense of the partiality of our agential perspective on events, whilst at the same time acknowledging the systematicity in our thinking about time. The strategy I have been advocating throughout is that a conception of time as a system of particulars enables us to do this. Thought which concerns particulars is partial and idiosyncratic in that it is a contingent matter where one actually stands in the nexus of particulars, and which things one is related to. But it also involves systematicity in that knowledge of particulars involves exercising one's command of the general structure of the domain. The key to making sense of our relation to the temporal world is to see that our temporal nature, and our fundamental temporal orientation towards the future, is what anchors us to time, and puts us in a position to grasp the temporal framework as a whole. We can only grasp the framework by locating ourselves in it, embedding the events of our lives into the causal temporal order. But in doing so we must thereby recognise the framework—and the events which make up our lives—as extending into the future.

A grasp of one's causal structure in the past thus reflects back on, and enlarges, one's understanding of oneself in the present. In reflecting on the past we must realise that the present cannot be fundamentally unlike the past; the same kinds of things are in progress now which we encounter as over and done with when we cast our eyes backwards. Yet there is a sense in which we can never properly grasp this; the present is never fully available to us until it is no longer present. We live in a kind of darkness, the Owl of Minerva spreading her wings only once the action is over. This is the genuine

insight contained in the claim that events exist only once they are over. But we can only appreciate its significance once we have abandoned the incoherent idea of a thing of the past.

BIBLIOGRAPHY

- Ahmed, Arif (2010). 'Causation and Decision'. In: *Proceedings of the Aristotelian Society* 110.2pt2, pp. 111–131. [See p. 103.]
- (2018a). *Newcomb's Problem*. Cambridge University Press. [See p. 85.]
- (2018b). 'The 'Why Ain'cha Rich?' Argument'. In: *Newcomb's Problem*. Cambridge University Press. [See p. 103.]
- Ahmed, Arif and Huw Price (2012). 'Arntzenius on 'Why Ain'Cha Rich?' In: *Erkenntnis* 77.1, pp. 15–30. [See p. 103.]
- Albert, David Z. (2000). *Time and Chance*. Harvard University Press. [See pp. 91, 92.]
- Anscombe, G. E. M. (1957). *Intention*. Cambridge, MA; London: Harvard University Press. [See p. 135.]
- (2015). 'Existence and the Existential Quantifier'. In: *Logic, Truth and Meaning: Writings of G.E.M. Anscombe*. Ed. by Mary Geach and Luke Gormally. Imprint Academic. [See p. 225.]
- Antinucci, Francesco and Ruth Miller (1976). 'How children talk about what happened'. In: *Journal of Child Language* 3.02, pp. 167–189. [See pp. 157, 158.]
- Appiah, Anthony (1994). 'Race, Culture, Identity: Misunderstood Connections'. In: *Tanner Lectures on Human Values*. [See p. 39.]
- Aristotle, Reeve (trans.) (2016). *Metaphysics*. Ed. by C. D. C. Reeve. Hackett.
- Aristotle and Roger (trans.) Crisp (2000). *Nicomachean Ethics*. Ed. by Roger Crisp. Cambridge: Cambridge University Press. [See p. 81.]
- Arntzenius, Frank (2008). 'No Regrets, Or: Edith Piaf Revamps Decision Theory'. In: *Erkenntnis* 68.2, pp. 277–297. [See p. 119.]

- Bach, Emmon (1986). 'The Algebra of Events'. In: *Linguistics and Philosophy* 9.1, pp. 5–16. [See p. 217.]
- Baron, Marcia (1988). 'Remorse and Agent-Regret'. In: *Midwest Studies in Philosophy* 13.1, pp. 259–281. [See p. 127.]
- Beebe, Helen, Christopher Hitchcock and Huw Price, eds. (2017). *Making a Difference: Essays on the Philosophy of Causation*. Oxford: Oxford University Press.
- Beere, Jonathan (2012). *Doing and Being: An Interpretation of Aristotle's Metaphysics Theta*. Oxford University Press UK. [See p. 225.]
- Belnap, Nuel and Mitchell Green (1994). 'Indeterminism and the Thin Red Line'. In: *Philosophical Perspectives* 8, pp. 365–388. [See p. 55.]
- Bigelow, John (1996). 'Presentism and Properties'. In: *Philosophical Perspectives* 10. Metaphysics, pp. 35–52. [See p. 69.]
- Black, Max (1955). 'Why Cannot an Effect Precede its Cause?'. In: *Analysis* 16.3, pp. 49–58. [See p. 88.]
- Bourne, Craig (2006). *A Future for Presentism*. Oxford University Press. [See p. 69.]
- Braddon-Mitchell, David (2013). 'Fighting the Zombie of the Growing Salami'. In: *Oxford Studies in Metaphysics* 8, pp. 351–362. [See p. 70.]
- Bratman, Michael (1987). *Intention, Plans, and Practical Reason*. Center for the Study of Language and Information. [See p. 144.]
- Broad, C. D. (1923). *Scientific Thought*. Routledge and Kegan Paul. [See p. 55.]
- Broadie, Sarah (1982). *Passage and Possibility: A Study of Aristotle's Modal Concepts*. Oxford University Press. [See p. 55.]
- Burge, Tyler (1974). 'Truth and Singular Terms'. In: *Noûs* 8.4, pp. 309–325. [See p. 223.]
- Button, Tim (2006). 'There's No Time Like the Present'. In: *Analysis* 66.2, pp. 130–135. [See p. 55.]
- Callender, Craig (2017). *What Makes Time Special?* Oxford University Press. [See pp. 53, 78.]
- Campbell, John (1993). 'A Simple View of Colour'. In: *Reality: Representation and Projection*. Ed. by John Haldane and Crispin Wright. Oxford: Oxford University Press. [See pp. 20, 48.]
- (1994). *Past, Space, and Self*. MIT Press. [See pp. 32, 75.]

-
- (1997). 'The Structure of Time in Autobiographical Memory'. In: *European Journal of Philosophy* 5.2, pp. 105–17. [See p. 191.]
 - (2006). 'Ordinary Thinking About Time'. In: *Time and History: Proceedings of the 28. International Ludwig Wittgenstein Symposium, Kirchberg Am Wechsel, Austria 2005*. Ed. by Michael Stöltzner and Friedrich Stadler. De Gruyter, pp. 1–12. [See pp. 171, 172.]
 - (2011). 'Why Do Language Use and Tool Use Both Count as Manifestations of Intelligence?' In: *Tool Use and Causal Cognition*. Ed. by Teresa McCormack, Christoph Hoerl and Stephen Butterfill. Oxford University Press. [See p. 173.]
 - Cartwright, Nancy (1979). 'Causal Laws and Effective Strategies'. In: *Noûs* 13.4, pp. 419–437. [See p. 82.]
 - Clarke, Thompson (1965). 'Seeing Surfaces and Physical Objects'. In: *Philosophy in America*. Ed. by Max Black. Allen and Unwin, pp. 98–114. [See p. 214.]
 - Clayton, Nicola S. and Anthony Dickinson (1998). 'Episodic-like memory during cache recovery by scrub jays'. In: *Nature* 395.6699, pp. 272–274. [See p. 161.]
 - Cocchiarella, Nino (1991). 'Quantification, Time, and Necessity'. In: *Philosophical Applications of Free Logic*. Ed. by Karel Lambert. Oxford University Press, pp. 242–256. [See p. 222.]
 - Collingwood, R. G. (1937). 'On the So-Called Idea of Causation'. In: *Proceedings of the Aristotelian Society* 38.1, pp. 85–112. [See p. 108.]
 - Correia, Fabrice and Sven Rosenkranz (2013). 'Living on the Brink, or Welcome Back, Growing Block!' In: *Oxford Studies in Metaphysics* 8, pp. 333–350. [See p. 70.]
 - (2018). *Nothing to Come*. Springer Verlag. [See pp. 55, 70.]
 - Cresswell, M. J. and G. E. Hughes (1984). *A Companion to Modal Logic*. New York: Methuen. [See pp. 203, 204.]
 - Crowther, Thomas (2011). 'The Matter of Events'. In: *The Review of Metaphysics*, pp. 3–39. [See p. 212.]
 - Dasgupta, Shamik (2009). 'Individuals: An Essay in Revisionary Metaphysics'. In: *Philosophical Studies* 145.1, pp. 35–67. [See p. 182.]
 - (2015). 'Substantivalism Versus Relationalism About Space in Classical Physics'. In: *Philosophy Compass* 10.9, pp. 601–624. [See p. 33.]

- Davidson, Donald (1980a). 'Actions, Reasons, and Causes'. In: *Essays on Actions and Events*. orig. 1963. Oxford: Clarendon Press. [See p. 179.]
- (1980b). *Essays on Actions and Events*. Oxford: Clarendon Press.
- (1980c). 'Events as Particulars'. In: *Essays on Actions and Events*. orig. 1970. Oxford: Clarendon Press. [See p. 216.]
- (1980d). 'The Logical Form of Action Sentences'. In: *Essays on Actions and Events*. orig. 1967. Oxford: Clarendon Press. [See p. 216.]
- (1984a). *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press.
- (1984b). 'Reality without Reference'. In: *Inquiries into Truth and Interpretation*. orig. 1977. Oxford: Oxford University Press. [See p. 219.]
- (1984c). 'The Inscrutability of Reference'. In: *Inquiries into Truth and Interpretation*. orig. 1979. Oxford: Oxford University Press. [See p. 219.]
- (1984d). 'Truth and Meaning'. In: *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press, pp. 17–36. [See p. 35.]
- Deasy, Daniel (2015). 'The Moving Spotlight Theory'. In: *Philosophical Studies* 172.8, pp. 2073–2089. [See p. 69.]
- Deng, Natalja (2013). 'On Explaining Why Time Seems to Pass'. In: *Southern Journal of Philosophy* 51.3, pp. 367–382. [See p. 55.]
- (2017). 'Making Sense of the Growing Block View'. In: *Philosophia* 45.3, pp. 1113–1127. [See p. 70.]
- Dougherty, Tom (2011). 'On Whether To Prefer Pain to Pass'. In: *Ethics* 121.3, pp. 521–537. [See p. 142.]
- Dowe, Phil (2000). *Physical Causation*. Cambridge University Press. [See pp. 89, 166.]
- Dummett, Michael (1978a). 'A Defense of McTaggart's Proof of the Unreality of Time'. In: *Truth and Other Enigmas*. orig. 1960. Harvard University Press. [See pp. 54, 55, 61.]
- (1978b). 'The Reality of the Past'. In: *Truth and Other Enigmas*. Duckworth, pp. 358–374. [See pp. 58, 207.]
- (1993). 'Causal Loops'. In: *The Seas of Language*. Oxford University Press. Chap. 15. [See pp. 88, 92.]

- Earman, John (2008). 'Reassessing the Prospects for a Growing Block Model of the Universe'. In: *International Studies in the Philosophy of Science* 22.2, pp. 135–164. [See pp. 55, 70.]
- Edgington, Dorothy (2011). 'Conditionals, Causation, and Decision'. In: *Analytic Philosophy* 52.2, pp. 75–87. [See p. 107.]
- (2014). 'Indicative Conditionals'. In: *Stanford Encyclopedia of Philosophy*. URL: <https://plato.stanford.edu/entries/conditionals/>. [See p. 85.]
- Eells, Ellery (1982). *Rational Decision and Causality*. Cambridge University Press. [See pp. 101, 103.]
- Egan, Andy (2007). 'Some Counterexamples to Causal Decision Theory'. In: *Philosophical Review* 116.1, pp. 93–114. [See p. 119.]
- Eilan, Naomi (1994). 'The First Person Perspective'. In: *Proceedings of the Aristotelian Society* 95.1, pp. 51–66. [See p. 48.]
- (1997). 'Objectivity and the Perspective of Consciousness'. In: *European Journal of Philosophy* 5.3, pp. 235–250. [See pp. 47, 62.]
- (2014). 'Intelligible Realism About Consciousness: A Response to Nagel's Paradox'. In: *Ratio* 27.1, pp. 32–52. [See p. 48.]
- Eisenberg, Ann R. (1985). 'Learning to describe past experiences in conversation'. In: *Discourse Processes* 8.2, pp. 177–204. [See p. 157.]
- Elster, Jon (2000). *Ulysses Unbound: Studies in Rationality, Precommitment, and Constraints*. Cambridge University Press. [See p. 144.]
- Evans, Gareth (1985a). 'Does Tense Logic Rest on a Mistake?' In: *Collected Papers: Gareth Evans*. Ed. by Gareth Evans. Oxford: Clarendon Press, pp. 343–363. [See p. 64.]
- (1985b). 'Understanding Demonstratives'. In: *Collected Papers*. Oxford: Clarendon Press. [See p. 61.]
- Evans, Gareth and John McDowell (1976a). 'Introduction'. In: *Truth and Meaning: Essays in Semantics*. Ed. by Gareth Evans and John McDowell. Oxford: Clarendon Press, pp. vii–xxiii. [See p. 35.]
- eds. (1976b). *Truth and Meaning: Essays in Semantics*. Oxford: Clarendon Press.
- Fernandes, Alison (2017). 'A Deliberative Approach to Causation'. In: *Philosophy and Phenomenological Research* 95.3, pp. 686–708. [See pp. 82, 83, 98, 120.]

- Fernandes, Alison (forthcoming). 'Freedom, Self-Prediction, and the Possibility of Time Travel'. In: *Philosophical Studies*, pp. 1–20. [See p. 96.]
- Fine, Kit (2005). 'Tense and Reality'. In: *Modality and Tense*. Oxford University Press. [See pp. 54, 64, 65.]
- (2008). 'In Defence of Three-Dimensionalism'. In: *Royal Institute of Philosophy Supplement* 83.62, pp. 1–16. [See p. 213.]
- Fivush, Robyn (2011). 'The Development of Autobiographical Memory'. In: *Annual Review of Psychology* 62.1, pp. 559–582. [See p. 157.]
- Fivush, Robyn, Judith Hudson and Katherine Nelson (1984). 'Children's long-term memory for a novel event: An exploratory study'. In: *Merrill-Palmer Quarterly*, pp. 303–316. [See pp. 156, 157.]
- Flew, Anothy (1954). 'Can an Effect Precede Its Cause?'. In: *Aristotelian Society Supplementary Volume* 28, pp. 27–62. [See p. 88.]
- Ford, Anton (2014). 'Action and Passion'. In: *Philosophical Topics* 42.1, pp. 13–42. [See p. 211.]
- Frederick, Shane, George Loewenstein and Ted O'Donoghue (2002). 'Time Discounting and Time Preference: A Critical Review'. In: *Journal of Economic Literature* 40.2, pp. 351–401. [See p. 142.]
- Gallistel, Charles R. (1993). *The Organization of Learning*. MIT Press. [See p. 161.]
- Galton, Antony (1984). *The Logic of Aspect: An Axiomatic Approach*. Clarendon Press. [See pp. 196, 230.]
- Galton, Antony and Riichiro Mizoguchi (2009). 'The Water Falls but the Waterfall Does Not Fall: New Perspectives on Objects, Processes and Events'. In: *Applied Ontology* 4.2, pp. 71–107. [See p. 212.]
- Garrett, Brian (2014). 'Black on Backwards Causation'. In: *Thought: A Journal of Philosophy* 3.3, pp. 230–233. [See p. 88.]
- (2016). 'Dummett on Bringing About the Past'. In: *Philosophia* 44.1, pp. 113–115. [See p. 89.]
- Geach, P. T. (1954). 'Form and Existence'. In: *Proceedings of the Aristotelian Society* 55.n/a, pp. 251–272. [See p. 225.]

- Gell, Alfred (1992). *The Anthropology of Time: Cultural Constructions of Temporal Maps and Images*. Oxford, Providence: Berg. [See p. 77.]
- Gibbard, Allan and William Harper (1978). 'Counterfactuals and Two Kinds of Expected Utility'. In: *Foundations and Applications of Decision Theory*. Ed. by A. Hooker, J. J. Leach and E. F. McClennen. D. Reidel, pp. 125–162. [See p. 102.]
- Gopnik, Alison, Clark Glymour et al. (2004). 'A Theory of Causal Learning in Children: Causal Maps and Bayes Nets'. In: *Psychological Review* 111.1, pp. 3–32. [See pp. 111, 166.]
- Gopnik, Alison and Tamar Kushnir (2003). 'Inferring hidden causes'. In: *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*. Ed. by R. Alterman and D. Kirsh. Boston. [See p. 111.]
- (2005). 'Children infer causal strength from probabilities and interventions'. In: *Psychological Science* 16.9, pp. 678–683. [See p. 111.]
- Gopnik, Alison, Tamar Kushnir, Chris Lucas et al. (2009). 'Inferring Hidden Causal Structure'. In: *Cognitive Science* 34.1, pp. 148–160. [See p. 111.]
- Gopnik, Alison, Tamar Kushnir and Laura Schulz (2007). 'Learning from Doing: Intervention and Causal Influence'. In: *Causal Learning: Psychology, Philosophy, and Computation*. Oxford University Press. [See p. 111.]
- Gopnik, Alison and Laura Schulz (2007). *Causal Learning: Psychology, Philosophy, and Computation*. Oxford University Press.
- Greene, Preston and Meghan Sullivan (2015). 'Against Time Bias'. In: *Ethics* 125.4, pp. 947–970. [See p. 142.]
- Hagmayer, York (2016). 'Causal Bayes Nets as Psychological Theories of Causal Reasoning: Evidence From Psychological Research'. In: *Synthese* 193.4, pp. 1107–1126. [See p. 111.]
- Halpern, Joseph Y. and Judea Pearl (2005). 'Causes and Explanations: A Structural-Model Approach. Part I: Causes'. In: *British Journal for the Philosophy of Science* 56.4, pp. 843–887. DOI: 10.1093/bjps/axi147. [See p. 115.]
- Hamond, Nina R. and Robyn Fivush (1991). 'Memories of Mickey Mouse: Young children recount their trip to Disneyworld'. In: *Cognitive Development* 6.4, pp. 433–448. [See p. 157.]

- Hare, Caspar (2007). 'Self-Bias, Time-Bias, and the Metaphysics of Self and Time.' In: *Journal of Philosophy* 104.7, pp. 350–373. [See p. 43.]
- Hawthorne, John and J. A. Cover (1998). 'A World of Universals.' In: *Philosophical Studies* 91, pp. 205–219. [See p. 187.]
- Hawthorne, John and Theodore Sider (2002). 'Locations.' In: *Philosophical Topics* 30.1, pp. 53–76. [See pp. 33, 187.]
- Higginbotham, James (2009). *Tense, Aspect and, Indexicality*. Oxford Studies in Theoretical Linguistics. Oxford, New York: Oxford University Press. [See p. 216.]
- Hilbert, David (2005). 'Color Constancy and the Complexity of Color.' In: *Philosophical Topics* 33.1, pp. 141–158. [See p. 39.]
- Hintikka, Jaakko (1962). *Knowledge and Belief*. Ithaca: Cornell University Press. [See p. 55.]
- Hitchcock, Christopher (2001). 'The Intransitivity of Causation Revealed in Equations and Graphs.' In: *Journal of Philosophy* 98.6, pp. 273–299. [See pp. 115, 117.]
- (2013). 'What is the 'Cause' in Causal Decision Theory?' In: *Erkenntnis* 78.1, pp. 129–146. [See p. 120.]
- (2017). 'Actual Causation.' In: *Making a Difference: Essays on the Philosophy of Causation*. Ed. by Helen Beebe, Christopher Hitchcock and Huw Price. Oxford: Oxford University Press. [See p. 120.]
- Hoerl, Christoph (2008). 'On Being Stuck in Time.' In: *Phenomenology and the Cognitive Sciences* 7.4, pp. 485–500. [See pp. 168, 170.]
- (2014). 'Do We (Seem to) Perceive Passage?' In: *Philosophical Explorations* 17.2, pp. 188–202. [See p. 55.]
- (2015a). 'Tense and the Psychology of Relief.' In: *Topoi* 34.1, pp. 217–231. [See p. 137.]
- (2015b). 'Writing on the Page of Consciousness.' In: *Proceedings of the Aristotelian Society* 115.3pt3, pp. 187–209. [See pp. 48, 50.]
- Hoerl, Christoph and Teresa McCormack (1999). 'Memory and Temporal Perspective: The Role of Temporal Frameworks in Memory Development.' In: *Developmental Review* 19, pp. 154–182. [See pp. 158, 159.]
- eds. (2001). *Time and Memory: Issues in Philosophy and Psychology*. Oxford University Press.

-
- (2011). ‘Tool Use, Planning, and Future Thinking in Children and Animals.’ In: *Tool Use and Causal Cognition*. Ed. by Teresa McCormack, Christoph Hoerl and Stephen Butterfill. Oxford University Press. [See p. 168.]
 - (2016). ‘Making Decisions About the Future: Regret and the Cognitive Function of Episodic Memory.’ In: *Seeing the future: Theoretical perspectives on future-oriented mental time travel*. Ed. by Kourken Michaelian, Stanley Klein and Karl Szpunar. Oxford University Press, pp. 241–266. [See p. 151.]
 - (2017). ‘The Development of Temporal Concepts: Learning to Locate Events in Time.’ In: *Timing and Time Perception* 5.3-4, pp. 297–327. [See p. 158.]
 - Hornsby, Jennifer (2012). ‘Actions and Activity.’ In: *Philosophical Issues* 22.1, pp. 233–245. [See p. 212.]
 - (2013). ‘Basic Activity.’ In: *Aristotelian Society Supplementary Volume* 87.1, pp. 1–18. [See pp. 211, 212.]
 - Horwich, Paul (1985). ‘Decision Theory in Light of Newcomb’s Problem.’ In: *Philosophy of Science* 52.3, pp. 431–450. [See pp. 92, 103.]
 - (1989). *Asymmetries in Time: Problems in the Philosophy of Science*. Bradford Books. [See p. 87.]
 - Hyman, John (2005). ‘What, If Anything, Are Colours Relative To?’ In: *Philosophy* 80.4, pp. 475–494. [See p. 38.]
 - Ismael, Jenann (2007). ‘Freedom, Compulsion, and Causation.’ In: *PSYCHE: An Interdisciplinary Journal of Research On Consciousness* 13. [See p. 154.]
 - Jackson, Frank (1982). ‘Epiphenomenal Qualia.’ In: *The Philosophical Quarterly* 32.127, p. 127. [See p. 42.]
 - Jeffrey, Richard (1981). ‘The Logic of Decision Defended.’ In: *Synthese* 48.3, pp. 473–492. [See pp. 103, 120.]
 - (1983). *The Logic of Decision*. University of Chicago Press. [See pp. 100, 141.]
 - Joyce, James M. (1999). *The Foundations of Causal Decision Theory*. Cambridge University Press. [See pp. 96, 101, 104, 107, 141, 149.]
 - (2000). ‘Why We Still Need the Logic of Decision.’ In: *Philosophy of Science* 67.3, p. 13. [See pp. 101, 141.]

- Joyce, James M. (2007). 'Are Newcomb Problems Really Decisions?' In: *Synthese* 156.3, pp. 537–562. DOI: 10.1007/s11229-006-9137-6. [See p. 103.]
- (2012). 'Regret and Instability in Causal Decision Theory'. In: *Synthese* 187.1, pp. 123–145. [See p. 119.]
- (2018). 'Deliberation and Stability in Newcomb Problems and Pseudo-Newcomb Problems'. In: *Newcomb's Problem*. Cambridge University Press. [See p. 103.]
- Kant, Immanuel (1992a). 'Concerning the Ultimate Ground of Differentiation of Directions in Space'. In: *Theoretical Philosophy, 1755–1770*. Ed. by D. Walford and R. Meerbote. The Cambridge Edition of the Works of Immanuel Kant. Cambridge: Cambridge University Press. [See p. 62.]
- (1992b). 'On the Forms and Principles of the Sensible and Intelligible World [inaugural dissertation]'. In: *Theoretical Philosophy, 1755–1770*. Ed. by D. Walford and R. Meerbote. The Cambridge Edition of the Works of Immanuel Kant. Cambridge: Cambridge University Press. [See p. 62.]
- (1992c). *Theoretical Philosophy, 1755–1770*. Ed. by D. Walford and R. Meerbote. The Cambridge Edition of the Works of Immanuel Kant. Cambridge: Cambridge University Press.
- (2004). *Prolegomena to Any Future Metaphysics That Will Be Able to Come Forward as Science [with selections from the Critique of Pure Reason]*. Ed. by Gary Hatfield. Cambridge University Press. [See p. 62.]
- Kaplan, David (1989). 'Demonstratives'. In: *Themes From Kaplan*. Ed. by Joseph Almog, John Perry and Howard Wettstein. Oxford University Press, pp. 481–563. [See p. 59.]
- Kosman, Aryeh (2013). *The Activity of Being: An Essay on Aristotle's Ontology*. Harvard University Press. [See p. 225.]
- Larson, Richard and Gabriel Segal (1995). *Knowledge of Meaning: An Introduction to Semantic Theory*. MIT Press. [See p. 35.]
- Leibniz, Gottfried Wilhelm (1991). *Discourse on Metaphysics and Other Essays*. Hackett Publishing Company. [See p. 182.]
- Levi, Isaac (1993). 'Rationality, Prediction, and Autonomous Choice'. In: *Canadian Journal of Philosophy* 23.sup1, pp. 339–363. [See p. 96.]
- Lewis, David (1981). "Why Ain'cha Rich?" In: *Noûs* 15.3, pp. 377–380. [See p. 103.]

-
- (1983). ‘Attitude De Dicto and De Se’. In: *Philosophical Papers, Vol. 1*. New York, Oxford: Oxford University Press. [See p. 59.]
 - (1987a). ‘Causal Decision Theory’. In: *Philosophical Papers, Vol. 2*. orig. 1981. New York, Oxford: Oxford University Press. [See p. 101.]
 - (1987b). ‘Counterfactual Dependence and Time’s Arrow’. In: *Philosophical Papers, Vol. 2*. orig. 1979. New York, Oxford: Oxford University Press. [See p. 86.]
 - (1987c). ‘Paradoxes of Time Travel’. In: *Philosophical Papers, Vol. 2*. New York, Oxford: Oxford University Press. [See p. 94.]
 - (1987d). *Philosophical Papers, Vol. 2*. New York, Oxford: Oxford University Press.
 - Lipman, Martin (2016). ‘Perspectival Variance and Worldly Fragmentation’. In: *Australasian Journal of Philosophy* 94.1, pp. 42–57. [See p. 43.]
 - (2018). ‘A Passage Theory of Time’. In: *Oxford Studies in Metaphysics: Volume 11*. Ed. by Karen Bennett and Dean Zimmerman. Oxford University Press, pp. 95–122. [See p. 69.]
 - MacBride, Fraser (2005). ‘The Particular–Universal Distinction: A Dogma of Metaphysics?’ In: *Mind* 114.455, pp. 565–614. [See p. 181.]
 - MacFarlane, John (2003). ‘Future Contingents and Relative Truth’. In: *Philosophical Quarterly* 53.212, pp. 321–336. [See p. 55.]
 - MacIntyre, Alasdair (1981). *After Virtue: A Study in Moral Theory*. University of Notre Dame Press. [See p. 133.]
 - Mackie, J. L. (1966). ‘The Direction of Causation’. In: *Philosophical Review* 75.4, pp. 441–466. [See p. 88.]
 - Marcus, Eric (2012). *Rational Causation*. Harvard University Press. [See p. 211.]
 - Markosian, Ned (2004). ‘A Defence of Presentism’. In: *Oxford Studies in Metaphysics* 1.3, pp. 47–82. [See p. 69.]
 - Martin, M. G. F. (2001). ‘Out of the Past: Episodic Recall as Retained Acquaintance’. In: *Time and Memory: Issues in Philosophy and Psychology*. Ed. by Christoph Hoerl and Teresa McCormack. Oxford University Press, pp. 257–284. [See p. 175.]
 - (2002). ‘Particular Thoughts and Singular Thought’. In: *Royal Institute of Philosophy Supplement* 51, pp. 173–214. [See p. 189.]
 - (2017). ‘Elusive Objects’. In: *Topoi* 36.2, pp. 247–271. [See p. 215.]

- Martin, M. G. F. (Unpublished). 'Desire in Time'. [See p. 135.]
- Maudlin, Tim (2012). *Philosophy of Physics: Space and Time*. Princeton University Press. [See p. 34.]
- McCall, Storrs (1976). 'Objective Time Flow'. In: *Philosophy of Science* 43.3, pp. 337–362. [See p. 55.]
- (1994). *A Model of the Universe*. Clarendon Press. [See p. 55.]
- McColgan, Kerry L. and Teresa McCormack (2008). 'Searching and planning: Young children's reasoning about past and future event sequences'. In: *Child Development* 79.5, pp. 1477–1497. [See p. 168.]
- McCormack, Teresa (2001). 'Attributing Episodic Memory to Animals and Children'. In: *Time and Memory: Issues in Philosophy and Psychology*. Ed. by Christoph Hoerl and Teresa McCormack. Oxford University Press. [See p. 161.]
- (2014). 'Three types of temporal perspective: Characterizing developmental changes in temporal thought: Development of temporal thought'. In: *Annals of the New York Academy of Sciences* 1326.1, pp. 82–89. [See p. 161.]
- McCormack, Teresa and Christoph Hoerl (2005). 'Children's Reasoning About the Causal Significance of the Temporal Order of Events'. In: *Developmental Psychology* 41.1, pp. 54–63. [See p. 169.]
- (2007). 'Young children's reasoning about the order of past events'. In: *Journal of Experimental Child Psychology* 98.3, pp. 168–183. [See p. 169.]
- (Forthcoming). 'Thinking in and About Time: A Dual Systems Perspective on Temporal Cognition'. In: *Behavioral and Brain Sciences*. [See pp. 76, 161, 164.]
- McCormack, Teresa, Christoph Hoerl and Stephen Butterfill (2011). *Tool Use and Causal Cognition*. Oxford University Press.
- McDowell, John (2011). 'Some Remarks on Intention in Action'. In: *The Amherst Lecture in Philosophy* 6. URL: http://www.amherstlecture.org/mcdowell12011/mcdowell12011_ALP.pdf (accessed on 05/05/2015). [See p. 211.]
- McTaggart, J. M. E. (1900). *The Nature of Existence*. Cambridge University Press. [See p. 54.]
- (1908). 'The Unreality of Time'. In: *Mind* 17.68, pp. 457–474. [See p. 54.]

- Meek, Christopher and Clark Glymour (1994). 'Conditioning and Intervening'. In: *British Journal for the Philosophy of Science* 45.4, pp. 1001–1021. [See p. 115.]
- Mellor, D. H. (1981). *Real Time*. Cambridge University Press. [See pp. 64, 196.]
- (1998). *Real Time II*. Routledge. [See pp. 64, 90.]
- Merlo, Giovanni (2016). 'Subjectivism and the Mental'. In: *Dialectica* 70.3, pp. 311–342. DOI: 10.1111/1746-8361.12153. [See p. 43.]
- Montague, Richard (1973). 'The Proper Treatment of Quantification in Ordinary English'. In: *Approaches to Natural Language*. Ed. by Patrick Suppes, Julius Moravcsik and Jaakko Hintikka. Dordrecht, pp. 221–242. [See p. 195.]
- Moore, A. W. (1997). *Points of View*. Clarendon Press. [See pp. 28, 37.]
- (2001). 'Apperception and the Unreality of Tense'. In: *Time and Memory: Issues in Philosophy and Psychology*. Ed. by Christoph Hoerl and Teresa McCormack. Oxford University Press. [See p. 65.]
- Moore, G. E. (1936). 'Is Existence a Predicate?' In: *Aristotelian Society Supplementary Volume* 15.1, pp. 154–188. [See p. 215.]
- Mourelatos, Alexander P. D. (1978). 'Events, processes, and states'. In: *Linguistics and Philosophy* 2.3, pp. 415–434. [See p. 217.]
- Munoz-Dardé, Véronique (2016). 'Puzzles of Regret'. In: *Philosophy and Phenomenological Research* 92.3, pp. 778–784. [See p. 135.]
- Nagel, Thomas (1979a). *Mortal Questions*. Cambridge University Press.
- (1979b). 'Subjective and Objective'. In: *Mortal Questions*. Cambridge University Press. [See p. 43.]
- (1979c). 'What Is It Like to Be a Bat?'. In: *Mortal Questions*. orig. 1974. Cambridge University Press. [See pp. 42, 74.]
- (1986). *The View From Nowhere*. Oxford University Press. [See pp. 23, 43, 64.]
- Neale, Stephen (2001). *Facing Facts*. Clarendon Press. [See p. 35.]
- Nelson, Katherine, ed. (1986). *Event Knowledge: Structure and Function in Development*. Hillsdale, NJ: Lawrence Erlbaum Associates. [See p. 156.]
- (1988). 'The Ontogeny of Memory for Real Events'. In: *Remembering Reconsidered*. Ed. by Ulric Neisser and Eugene Winograd. Cambridge: Cambridge University Press, pp. 244–276. [See p. 156.]

- Nelson, Katherine (1996). *Language in Cognitive Development: The Emergence of the Mediated Mind*. Cambridge University Press. [See p. 156.]
- Newton-Smith, W. (1980). *The Structure of Time*. Routledge & Kegan Paul. [See p. 196.]
- Nozick, Robert (1969). 'Newcomb's Problem and Two Principles of Choice'. In: *Essays in Honor of Carl G. Hempel*. Ed. by Nicholas Rescher. Reidel, pp. 114–146. [See p. 97.]
- O'Connor, Eimear, Teresa McCormack, Sarah R. Beck et al. (2015). 'Regret and adaptive decision making in young children'. In: *Journal of Experimental Child Psychology* 135, pp. 86–92. [See p. 151.]
- O'Connor, Eimear, Teresa McCormack and Aidan Feeney (2014). 'Do Children Who Experience Regret Make Better Decisions? A Developmental Study of the Behavioral Consequences of Regret'. In: *Child Development*, pp. 1995–2010. [See p. 151.]
- Ogihara, Toshiyuki (2007). 'Tense and Aspect in Truth-Conditional Semantics'. In: *Lingua* 117.2, pp. 392–418. [See pp. 58, 195.]
- O'Shaughnessy, Brian (1971). 'Processes'. In: *Proceedings of the Aristotelian Society* 72, pp. 215–240. [See p. 228.]
- Parfit, Derek (1984). *Reasons and Persons*. Oxford University Press. [See pp. 142, 152.]
- Parsons, Terence (1990). *Events in the Semantics of English: A Study in Subatomic Semantics*. The MIT Press. [See p. 216.]
- Partee, Barbara Hall (1973). 'Some Structural Analogies Between Tenses and Pronouns in English'. In: *Journal of Philosophy* 70.18, pp. 601–609. [See p. 58.]
- Paul, L. A. (2010). 'Temporal Experience'. In: *Journal of Philosophy* 107.7, pp. 333–359. [See p. 54.]
- Peacocke, Christopher (2011). 'Representing Causality'. In: *Tool Use and Causal Cognition*. Ed. by Teresa McCormack, Christoph Hoerl and Stephen Butterfill. Oxford University Press. [See p. 173.]
- Pearl, Judea (2009). *Causality: Models, Reasoning and Inference*. 2nd ed. orig. 2000. Cambridge University Press. [See pp. 108, 113.]
- Percival, Philip (2002). 'A Presentist's Refutation of Mellor's McTaggart'. In: *Royal Institute of Philosophy Supplement* 50, pp. 91–. DOI: 10.1017/S1358246100010523. [See pp. 64, 68.]

- Perry, John (1979). 'The Problem of the Essential Indexical'. In: *Noûs* 13.1, pp. 3–21. [See p. 59.]
- Pink, Thomas (1996). *The Psychology of Freedom*. Cambridge University Press. [See p. 144.]
- Popper, Karl (1956). 'The Arrow of Time'. In: *Nature* 177.4507, pp. 538–538. [See p. 91.]
- Povinelli, Daniel J. et al. (1999). 'Development of young children's understanding that the past is causally bound to the present'. In: *Developmental Psychology* 35.6, pp. 1426–1439. [See p. 169.]
- Price, Huw (1986). 'Against Causal Decision Theory'. In: *Synthese* 67.2, pp. 195–212. [See pp. 92, 103.]
- (1991). 'Agency and Probabilistic Causality'. In: *British Journal for the Philosophy of Science* 42.2, pp. 157–176. [See pp. 82, 107.]
- (1997). *Time's Arrow and Archimedes' Point: New Directions for the Physics of Time*. Oxford University Press USA. [See p. 83.]
- (2012). 'Causation, Chance, and the Rational Significance of Supernatural Evidence'. In: *Philosophical Review* 121.4, pp. 483–538. [See p. 99.]
- (2017). 'Causation, Intervention and Agency—Woodward on Menzies and Price'. In: *Making a Difference: Essays on the Philosophy of Causation*. Ed. by Helen Beebe, Christopher Hitchcock and Huw Price. Oxford: Oxford University Press. [See pp. 112, 114.]
- Price, Huw and Peter Menzies (1993). 'Causation as a Secondary Quality'. In: *British Journal for the Philosophy of Science* 44.2, pp. 187–203. [See pp. 82, 114.]
- Prior, A. N. (1959). 'Thank Goodness That's Over'. In: *Philosophy* 34.128, pp. 12–17. [See p. 136.]
- (1967). *Past, Present and Future*. Clarendon Press. [See pp. 55, 196, 205.]
- (1972). 'The Notion of the Present'. In: *The Study of Time*. Ed. by J. T. Fraser, F. Haber and G. Muller. Springer Verlag, pp. 320–323. [See p. 67.]
- (2003). 'Changes in Events and Changes in Things'. In: *Papers on Time and Tense*. 2nd ed. orig. 1962. Oxford University Press. [See pp. 67, 194, 230.]
- Quine, W. V. O. (1948). 'On What There Is'. In: *Review of Metaphysics* 2.1, pp. 21–38. [See p. 215.]

- Ramsey, F. P. (1978). 'General Propositions and Causality'. In: *Foundations: Essays in Philosophy, Logic, Mathematics and Economics*. Ed. by D. H. Mellor. Humanities Press; Routledge. [See p. 154.]
- Rassool, Sara B. and Pieter W. Nel (2012). 'Experiences of Causing an Accidental Death: An Interpretative Phenomenological Analysis'. In: *Death Studies* 36.9, pp. 832–857. [See p. 129.]
- Reese, Elaine (2002). 'A model of the origins of autobiographical memory'. In: *Progress in Infancy Research*. Ed. by H. Hayne J. W. Hagen. Vol. 2. Mahwah, NJ: Erlbaum. [See p. 157.]
- Reichenbach, Hans (1956). *The Direction of Time*. Dover Publications. [See pp. 90, 91, 95, 186.]
- (1958). *The Philosophy of Space and Time*. orig. 1927; trans. Maria Reichenbach and John Freud, 1957. New York: Dover. [See p. 92.]
- (1980). *Elements of Symbolic Logic*. orig. 1947. Dover Publications. [See p. 158.]
- Rennick, Stephanie (2015). 'Things Mere Mortals Can Do, but Philosophers Can't'. In: *Analysis* 75.1, pp. 22–26. [See p. 96.]
- Roache, Rebecca (2009). 'Bilking the Bilking Argument'. In: *Analysis* 69.4, pp. 605–611. [See p. 88.]
- Rödl, Sebastian (2012). *Categories of the Temporal: An Inquiry Into the Forms of the Finite Understanding*. Harvard University Press. [See p. 229.]
- Rumfitt, Ian (1993). 'Content and Context: The Paratactic Theory Revisited and Revised'. In: *Mind* 102.407, pp. 429–454. DOI: 10.1093/mind/102.407.429. [See pp. 34, 60.]
- (2010). 'Logical Necessity'. In: *Modality: Metaphysics, Logic, and Epistemology*. Ed. by Bob Hale and Aviv Hoffmann. Oxford University Press. [See p. 218.]
- Russell, Bertrand (1912). 'On the Notion of Cause'. In: *Proceedings of the Aristotelian Society* 7, pp. 1–26. [See pp. 82, 91.]
- (1940). *The Philosophy of Logical Atomism*. orig. 1918–19. Routledge. [See p. 215.]
- (1956). 'On the Relations of Universals and Particulars'. In: *Logic and Knowledge: Essays, 1901–1950*. orig. 1911. Macmillan. [See p. 183.]
- (2001 (orig. 1912)). *The Problems of Philosophy*. Oxford: Oxford University Press. [See p. 75.]

- Salmon, Wesley (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton University Press. [See pp. 89, 166.]
- Savage, Lenoard J. (1951). 'The Theory of Statistical Decision'. In: *Journal of the American Statistical Association* 46, pp. 55–67. [See p. 126.]
- (1954). *The Foundations of Statistics*. Wiley Publications in Statistics. [See p. 100.]
- Schank, Roger C. and Robert P. Abelson (1977). *Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures*. Oxford: Laurence Erlbaum. [See p. 156.]
- Shoemaker, Sydney (2003). 'Causality and Properties'. In: *Identity, Cause, and Mind*. orig. 1980. Oxford; New York: Oxford University Press. [See p. 173.]
- Shultz, Thomas R. (1982). 'Rules of Causal Attribution'. In: *Monographs of the Society for Research in Child Development* 47.1, pp. 1–51. [See p. 166.]
- Sider, Theodore (2001). *Four Dimensionalism: An Ontology of Persistence and Time*. Oxford University Press. [See p. 70.]
- Skow, Bradford (2015). *Objective Becoming*. Oxford University Press UK. [See pp. 55, 62, 69.]
- Sloman, Steven (2005). *Causal Models: How People Think About the World and its Alternatives*. Oup Usa. [See p. 111.]
- Sorabji, Richard (1988). *Matter, Space, and Motion: Theories in Antiquity and Their Sequel*. London: Duckworth. [See p. 192.]
- Soteriou, Matthew (2000). 'The Particularity of Visual Perception'. In: *European Journal of Philosophy* 8.2, pp. 173–189. [See p. 189.]
- (2013). *The Mind's Construction: The Ontology of Mind and Mental Action*. Oxford University Press. [See pp. 76, 145.]
- (2019). 'Waking Up and Being Conscious'. In: *Aristotelian Society Supplementary Volume* 93.1, pp. 111–136. [See p. 138.]
- Spirtes, Peter et al. (2000). *Causation, Prediction, and Search*. 2nd ed. orig. 1993. Mit Press: Cambridge. [See p. 108.]
- Steenhagen, Maarten (2017). 'False Reflections'. In: *Philosophical Studies* 174.5, pp. 1227–1242. [See p. 188.]

- Steward, Helen (2012). 'Actions as Processes'. In: *Philosophical Perspectives* 26.1, pp. 373–388. [See pp. 212, 228.]
- (2013). 'Processes, Continuants, and Individuals'. In: *Mind* 122.487, pp. 781–812. [See p. 212.]
- Stout, Rowland (1997). 'Processes'. In: *Philosophy* 72.279, pp. 19–27. [See p. 212.]
- (2003). 'The Life of a Process'. In: *Process Pragmatism*. Ed. by Guy Debrock. Rodopi. [See p. 212.]
- (2016). 'The Category of Occurrent Continuants'. In: *Mind* 125.497, pp. 41–62. [See p. 212.]
- (2018). *Process, Action, and Experience*. Oxford University Press. [See p. 212.]
- Strawson, P. F. (1959). *Individuals: An Essay in Descriptive Metaphysics*. London, New York: Routledge. [See p. 34.]
- (2006). *The Bounds of Sense: An Essay on Kant's Critique of Pure Reason*. orig. 1966. London, New York: Routledge. [See p. 45.]
- Stroud, Barry (2000). *The Quest for Reality: Subjectivism and the Metaphysics of Colour*. Oxford University Press. [See pp. 22, 41.]
- Suddendorf, Thomas and Janie Busby (2003). 'Mental time travel in animals?' In: *Trends in Cognitive Sciences* 7.9, pp. 391–396. [See p. 161.]
- Sullivan, Meghan (2012a). 'Problems for Temporary Existence in Tense Logic'. In: *Philosophy Compass* 7.1, pp. 43–57. [See p. 222.]
- (2012b). 'The Minimal A-Theory'. In: *Philosophical Studies* 158.2, pp. 149–174. [See p. 69.]
- Taylor, Barry (1985). *Modes of Occurrence: Verbs, Adverbs and Events*. Vol. 2. Aristotelian Society Series. Oxford: Blackwell. [See p. 216.]
- Taylor, Charles (1989). *Sources of the Self: The Making of the Modern Identity*. Harvard University Press. [See pp. 133, 153.]
- Thomason, Richmond H. (1970). 'Indeterminist Time and Truth-Value Gaps'. In: *Theoria* 36.3, pp. 264–281. [See p. 55.]
- Thompson, Michael (2008). *Life and Action: Elementary Structures of Practice and Practical Thought*. Cambridge, MA: Harvard University Press. [See pp. 135, 211.]

- Tooley, Michael (1997). *Time, Tense, and Causation*. Oxford University Press. [See pp. 55, 70.]
- Van Fraassen, Bas C. (1995). 'Belief and the Problem of Ulysses and the Sirens'. In: *Philosophical Studies* 77.1, pp. 7–37. DOI: 10.1007/bf00996309. [See p. 120.]
- Vendler, Zeno (1957). 'Verbs and Times'. In: *The Philosophical Review*, pp. 143–160. [See p. 158.]
- Wagner, Laura (2001). 'Aspectual influences on early tense comprehension'. In: *Journal of Child Language* 28.03. [See p. 157.]
- (2012). 'Primary Language Acquisition'. In: *The Oxford Handbook of Tense and Aspect*. Ed. by Robert I. Binnick. Oxford Handbooks. Oxford University Press. [See p. 157.]
- Wallace, R. Jay (2013). *The View From Here: On Affirmation, Attachment, and the Limits of Regret*. Oup Usa. [See p. 129.]
- Weist, Richard M. (1986). 'Tense and Aspect'. In: *Language Acquisition: Studies in First Language Development*. Ed. by P. Fletcher and M. Garman. 2nd ed. Cambridge: Cambridge University Press, pp. 356–374. [See p. 158.]
- Wiggins, David (2001). *Sameness and Substance Renewed*. Cambridge University Press. [See pp. 190, 225.]
- Williams, Bernard (1978). *Descartes: The Project of Pure Enquiry*. Routledge. [See pp. 22, 26.]
- (1981). 'Moral Luck'. In: *Moral Luck: Philosophical Papers 1973–1980*. orig. 1976. Cambridge University Press. Chap. 2. [See p. 125.]
- (1985). *Ethics and the Limits of Philosophy*. Harvard University Press. [See p. 26.]
- (1992). *Shame and Necessity*. University of California Press. [See p. 96.]
- (1995a). 'Internal Reasons and the Obscurity of Blame'. In: *Making Sense of Humanity: And Other Philosophical Papers 1982–1993*. Cambridge University Press. [See p. 133.]
- (1995b). *Making Sense of Humanity: And Other Philosophical Papers 1982–1993*. Cambridge University Press.

- Williams, Bernard (1995c). 'Voluntary Acts and Responsibility Agents'. In: *Making Sense of Humanity: And Other Philosophical Papers 1982–1993*. orig. 1987. Cambridge University Press. [See p. 133.]
- (2009). 'Life as Narrative'. In: *European Journal of Philosophy* 17.2, pp. 305–314. [See p. 133.]
- Williamson, Timothy (2002). 'Necessary Existents'. In: *Royal Institute of Philosophy Supplement*. Ed. by A. O'Hear. Cambridge University Press, pp. 269–87. [See p. 221.]
- Wilson, George M. (1989). *The Intentionality of Human Action*. Stanford University Press. [See p. 216.]
- Wollheim, Richard (1974). 'Needs, Desires and Moral Turpitude: Richard Wollheim'. In: *Royal Institute of Philosophy Supplement* 8, pp. 162–179. [See p. 135.]
- Woods, Michael (1976). 'Existence and Tense'. In: *Truth and Meaning*. Ed. by Gareth Evans and John McDowell. Oxford: Clarendon Press. [See p. 225.]
- Woodward, James (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press. [See pp. 83, 87, 108, 110, 117.]
- (2007). 'Interventionist Theories of Causation in Psychological Perspective'. In: *Causal Learning: Psychology, Philosophy, and Computation*. Oxford University Press. [See pp. 115, 172, 173.]
- (2011). 'A Philosopher Looks at Tool Use and Causal Understanding'. In: *Tool Use and Causal Cognition*. Ed. by Teresa McCormack, Christoph Hoerl and Stephen Butterfill. Oxford University Press. [See p. 173.]