

Dissertation Proposal

Understanding Causation through Deliberation:

Using an Epistemic Model of Deliberation to Account for Features of Causation

Alison Fernandes

Columbia University

19 July 2013

My aim in this dissertation is to show how an epistemic model of decision-making contributes to our understanding of causation, and why we deliberate towards and control the future but not the past. This proposal begins with a basic summary of the chapters before going on to discuss their contents in more detail. An annotated bibliography follows, arranged thematically in rough accord with the chapter headings.

Table of Contents

1. Introduction

The introduction summarises the project and considers two main motivations for the project. Firstly, I argue that causation has no clear place in our best understanding of fundamental physics, a set of dynamical laws that relate global states of a system across time. This leaves a puzzle concerning how causation will fit into our understanding of the world. Secondly, I argue that there are general

concerns about fitting agency into the world as understood by science. This manifests in the freedom and determinism debate as well as more general concerns about the relation between theoretical and practical reason—concerns I suggest are well-encapsulated in the work of the German Idealism.

2. An Epistemic Model of Deliberation

The second chapter presents my own account of how apparent freedom in deliberation can be understood in terms of what beliefs an agent has. Drawing on work by Levi, I argue for a set of minimal epistemic conditions on deliberation, and show how these can partly account for apparent free choice in deliberative contexts, without merely presupposing a primitive freedom to intervene in the world. I then consider what consequences such an account has for the relation between theoretical and practical reason.

3. Defending the Model

The third chapter compares the above account to a number of close rivals that also attempt to explain the appearance of freedom in deliberation. In particular, I look at accounts by Fichte, Joyce and Velleman, in order to defend and bring out the advantages of my own view. I argue that even though Velleman's account is labelled as 'epistemic', it crucially presupposes a primitive interventionist freedom—similar in type to Fichte's. This creates internal problems for Joyce and

Velleman's account, making it unsuitable for understanding how such apparent freedom arises, and for understanding interventionist notions. I also consider how Fichte's account of the relation between theoretical and practical reason differs from my own.

4. Understanding Causation through Deliberation

Having developed a plausible epistemic model of deliberation, I now consider how this can be used to further our understanding of causation. Drawing on work by Price, Ismael and others, I suggest in what way causal structure can be understood by reference to an epistemic accounts of decision-making. I show that appealing to evidential relations to elucidate causal structure requires counterfactuals, and a notion of intervention based in free deliberation. I also argue that an important part of the justification of the counterfactual algorithm used appeals to pragmatics—what concepts will be useful to agents—an aspect that an account focussed on decision-making and deliberation is well-placed to handle.

5. Asymmetries in Control and Causation

I end by considering how this model of deliberation also finds application in explaining the future-directedness of decision-making and control, drawing on work by Loewer and Albert. I argue that we gain a deeper understanding of the

asymmetry of control by appealing to an epistemic model of deliberation, rather than assuming a primitive type of decision-intervention. I also consider to what degree we can explain an asymmetry in causation itself by appeal to such models.

6. Conclusion

I conclude with a summary of major points, including whether this epistemic account of deliberation and its relation to causal structure answers to broader concerns present in Fichte's work (concerning the relation of theoretical and practical reason) and what type of reduction in causation has been achieved.

Directions for further research are also suggested.

General Model

Drawing on work in action and decision theory, I develop an epistemic account of free deliberation—a model in which an agent's sense of free choice in deliberation is explained in terms of her epistemic states, and not in terms of a more metaphysical kind of contra-causal freedom. This epistemic model offers a plausible model for how apparently free action is possible in a world with deterministic laws. And it helps explicate a minimal notion of intervention in the present, such that different future states may be brought about by free choices the agent makes in the present. This model is presented in chapter 2 and defended further in chapter 3 and throughout the dissertation.

Motivations

The project has two main motivations, presented and defended in chapter 1 and revisited in later chapters. The first is to understand how the phenomenon of agency fits into the world as understood by natural science: a world with causes that prompt action and decision, and with statistical or deterministic laws that fix the occurrence of events at all other times, given the state of the world at any given time. Concern with this problem in the 20th century prompted ongoing debates over the compatibility of freedom and determinism. But the problem has much deeper routes, amongst others, in the German Idealist tradition. Concern with this broader problem will lead to look at work by J. G. Fichte, and his attempt to extend the Kantian project and ground theoretical reason in practical reason—why it is that such a project is reasonable and what is required to fulfil it.

This epistemic model of deliberation partly answers to the first motivation by offering an account of the relation between theoretical and practical reason. The model implies that while theoretical reason needs to leave room for the appearance of practical free choice, theoretical reason is also essential in various ways to the appearance and operation of practical reason. In considering the interaction of these two spheres, I draw on the work of Fichte, comparing and contrasting his account to my own (chapter 3). Fichte attempted to partially ground theoretical reason in practical reason, and explain how we seem to have unmediated control over our bodily actions. I consider how Fichte accounts for these practical kinds of interventions, comparing his account to my own, and considering

whether he can successfully account for the interactions between practical and theoretical reason. While Fichte's direct influence on contemporary analytic accounts of agency is slight, similar views that also ground practical reason in spontaneous activities of the mind are still prominent in contemporary literature (such as in the Kant-inspired work of Velleman).

A second motivation for the project (also introduced in chapter 1) is to understand how causal structure fits into the world, given that the fundamental dynamics of the world are not causal. The fundamental dynamical laws posited in contemporary physics relate time slices of the world to one another, not individual causes to effects. Not do laws have the type of past–future asymmetry that is taken to be characteristic of causation. This seems to make it mysterious how higher-level causal-laws can be built up out of these lower-level dynamics. In the second half of the dissertation, I will show how the epistemic model of deliberation helps with our understanding of causation, in ways I will elucidate below.

More on the Epistemic Model

While a number of philosophers have argued that decision-contexts are important for understanding causation, less work has been done in considering what models of deliberation are appropriate to such a project—theorists tend to blur or ignore the distinctions between the different types of accounts. I argue that an epistemic model grounded in ignorance conditions, amongst its other advantages, provides what is needed

here. Drawing on the work of Isaac Levi, in chapter 2, I claim that in order for an option to be seen as available to an agent during deliberation, the agent *cannot be certain* whether she will or will not choose that option, or that the option will obtain regardless of her choice, and *must be certain* that the choice of the option as good evidence for the option coming about. I argue that these requirements should be seen as something like conceptual truths concerning the nature of deliberation. The beliefs an agent has then delimit the decision-space available to her during deliberation, and, combined with an account of decision, help explain her sense of free choice—her sense that various options are available to her and her sense that these options are brought about by her decision.

I briefly consider other models of deliberation that are appealed to in order to ground causation in chapter 3. One is the type of model that Fichte seems to employ which I alluded to above—identifying the mind as a certain type of spontaneous practical activity that is then used to ground both its further practical and theoretical activities. But I also spend some time looking at models that attempt to account for free choice in virtue of the nature of decision. In particular I look at accounts by James M. Joyce and Jenann Ismael that identify decision as a belief about that very decision (decisions are self-constituting beliefs) and work by J. D. Velleman that identifies decision as a causally self-fulfilling belief about action. Joyce and Velleman use the latter's notion of 'epistemic freedom' to account for the appearance of free choice. These accounts are relevant not only because they provide competing models of free deliberation, but because they are used to argue against the very kind of epistemic limits that are employed in the model I prefer. I argue that these accounts in fact presuppose a primitive notion of intervention, not only making

them unsuitable for the larger project, but also creating internal problems for the accounts and their aspirations—problems that are obscured when they aren't properly distinguished from epistemic accounts of deliberation grounded in ignorance conditions.

Applying the Model

After introducing this epistemic model of free deliberation in the first half of the dissertation, I then consider how it is relevant for these debates in philosophy of science regarding causation. In particular, in chapter 4 I consider how the epistemic account of free deliberation can elucidate the concept of causation and to some extent help ground causal structure in more basic dynamical relations. A number of contemporary theorists approach the above problems regarding causation by appealing to counterfactuals and notions of intervention. These are reductivists view of causation, which attempt to explain causation, or provided a metaphysical grounding for causation, in terms of counterfactuals, and sometimes ultimately in non-causal lawlike relations. (These contrast with primitivist views of causation, which not only deny that the concept of causation can be analysed into more basic conceptual components, but claim causation requires no philosophical elucidation or explanation. For primitivists, we have relatively direct access to the causal or dispositional structure of the world, and no account in terms of non-causal entities is required to explain this access.)

But reductivist views of causation, such as David Lewis', and even broadly manipulationist view such as James Woodward's, face the problem of explaining how

counterfactual reasoning is relevant to causal relations that hold in the actual world and how notions of intervention are relevant. My project attempts to fill out such accounts by showing how various aspects of causation can be related to more basic dynamical law-like structures and basic evidential relations. In particular, by appealing to an epistemic model of deliberation, and the information that agents require in deliberation, I argue that we can explain how it is that interventions in the world seem possible and why it is that counterfactual reasoning is relevant in such cases. In order to understand the application of counterfactuals and notions of intervention, we need to consider free action and decision-making, and the kind of evidential connections that are important in decision-making. This gives us an account of how intervention appears possible and why the kind of information encoded in causal and counterfactual claims is relevant to our actual interactions with the world. Whether this implies that agency is part of grounding causation, or merely relevant to understanding the concept, important lacunae in reductivist views of causation are filled in by considering the epistemic states of deliberating agents more closely.

The relevance of deliberation to causation also provides a further response to the first motivation for the project: to understand how the phenomenon of agency fits into the world as understood by natural science. An epistemic account of deliberation answers to this broad concern by showing two things: firstly, how decision-making is compatible with the law-like and causal nature of the world and, secondly, by showing how the concept of causation is best understood by reference to decision-making. An epistemic account of deliberation is not only compatible with a world that operates according to

deterministic laws and following a causal order; in order to make fullest sense of what it is to have a causal order, and why a concept of causation is relevant for us, we need to consider the information that is relevant for agents in deliberation. And so understanding agency is bound up with understanding a very basic aspect of the world.

In chapters 4, I consider accounts by Huw Price and Jenann Ismael (amongst others) that attempt to ground causation, or to elucidate the concept of causation, in terms of agency and decision. The rough idea is that causal claims encode evidential information about what will occur, given possible interventions in the world, when the evidence of a deliberating agent is held fixed. In other words, rather than looking for causal structure in the fundamental law-like structure of the world, we should look for causal structure as what evidential connections are used by agents in deliberation—agents who hold fixed their evidence, and consider what further states their decisions will be evidence for. I question whether the evidential connections appealed to in such accounts can be understood independently of causation—and so whether such accounts are capable of providing reduction or merely conceptual elucidation. And I consider further limitations to such projects. While I am sceptical that such accounts can metaphysically ground the notion of causation that is employed widely in science, they do show important conceptual connections between deliberation, agency, causation and evidence.

Looking at decision-contexts also helps to highlight important pragmatic aspects of causation. In order to justify accounts of causation and counterfactuals, we need to show that such concepts will be useful to agents for getting around in the world—and so for

scientific theorising. While such an explanation *is* necessarily circular, involving appealing to counterfactuals or causation itself, it is important for a full understanding of the concept of causation. In chapter 4 I briefly sketch how such justifications should proceed—one should show that, given the account of counterfactuals or causation, these will be accessible relations that are usefully employed by agents in achieving goals and intentions. And, as above, I consider whether such circularity prevents such analyses of causation from being reductive.

A brief note: there is a debate in decision theory between evidential decision theory and causal decision theory—over whether it is rational to decide based on what your decisions are good evidence for, or what they are causes for. While relevant to understanding the conceptual connections between agency and causation, and while I relate my discussion to this debate, I do not consider intend to make an original contribution to this literature. The debate is often acknowledged to be dead-locked, with neither set of proponents expecting to sway the others with their arguments. I do state my own preference for evidential decision theory, since this provides greater scope for reductivist accounts of causation that appeal to rational deliberating. And I will show where the two types of theories play equivalent roles in the accounts. But I won't directly argue against others who favour causal decision theory.

Finally, in chapter 5, I suggest how such an epistemic model of deliberation is relevant for understanding the (perhaps only apparent) past–future asymmetry in causation: why it seems we cannot affect the past, but can affect the future, or, more controversially, why

effects necessarily come after their causes. Even if one is sceptical of the grander project that attempts to ground causation in evidential relations, one may still think that the asymmetry of causation may be explained in this way. Here I will be looking to accounts by Barry Loewer and David Albert, amongst others, that use statistical mechanics to explain the direction of causation.

The move to these statistical mechanical accounts is partly motivated by problems with Lewis' account, gaps left in Woodward's, and a resistance to seeking a causal asymmetry in the fundamental laws, or in a temporal asymmetry intrinsic to agency. I agree with these motivations. But I argue that decision contexts still need to be appealed to in these statistical mechanical explanations. These explanations are incomplete without appealing to a serial structure of agency, which is then combined with asymmetric but contingent features of the world to explain the overall asymmetry. More precisely, I appeal firstly to the epistemic model of deliberation and to the fact that in paradigmatic cases, deliberative action will follow a serial ordering of deliberation–decision–action. And I appeal to an initial low entropy condition of the universe, the so-called 'Past Hypothesis', which is used to explain the asymmetric second law of thermodynamics. I argue that these two features combined imply that agency will run past–future in our world. If there is indeed a close conceptual connection between agency and causation, then this asymmetry in agential deliberation and action can be parlayed into an asymmetry in causation. Or at least that is the intention of some such accounts. I will suggest that again there are limitations in what such accounts can be used to show; they are limited to cases of causation where the notion of intervention makes sense, and it is not necessarily clear

how to apply notions of intervention to all aspects of the fundamental level or to system wide events. But such accounts remain relevant to causation in the higher-level sciences.

Overall Aims

- 1) To develop and argue for an epistemic model of free deliberation, that explains the sense of free choice in deliberation in terms of both an agent's epistemic states, and an available decision space. While I cannot defend this account against the full range of alternatives, I will consider its advantages and weaknesses in comparison with close rivals, and consider how it is relevant in explanatory projects in the philosophy of science.

- 2) To provide further support for broadly reductivist views of causation, by showing how they can effectively appeal to structural and evidential features of deliberation to explain or elucidate important parts of our concept of causation. While I will note some limitations to the type of reduction that can be achieved, I intend to give the reductivist project strong support in developing and applying an epistemic model of free deliberation.

- 3) To show how work from diverse fields including history of philosophy, decision theory, theory of action, philosophy of science and physics is relevant to a core set of metaphysical questions concerning agency and causation. In developing an

account of my own, I aim to facilitate discussion between these fields on the nature of intervention and agency-based views of causation.

Annotated Bibliography

The following bibliography refers to a selection of major works I will be using.

Additional works by the same author, or further critical work, have often been omitted.

1. Fichte (chapters 1 and 3)

Fichte explains the relation between practical and theoretical reason, and how freedom is possibly in a causally determined world, in the following two early presentations, known as part of the Jena *Wissenschaftslehre*. I intend to focus on the second, with reference to the first.

Fichte, J. G. 1794–5. *Grundlage der gesamten Wissenschaftslehre (Foundations of the Entire Science of Knowledge)*, trans. Peter Heath. in *Fichte: Science of Knowledge (Wissenschaftslehre)*, ed. Peter Heath and John Lachs, New York: Appleton-Century-Crofts, 1970.

———. 1796–9. *Wissenschaftslehre nova methodo (Foundations of Transcendental Philosophy)*, trans. and ed. Daniel Breazeale, Ithaca: Cornell University Press, 1992.

Secondary literature:

Breazeale, Daniel and Tom Rockmore. 2002 eds. *New Essays on Fichte's Later Jena Wissenschaftslehre*, Evanston, IL: Northwestern University Press.

Breazeale, Daniel. 1996. The Theory of Practice and the Practice of Theory: Fichte and the "Primacy of Practical Reason". *International Philosophical Quarterly* 36(1): 47-64.

2. Epistemic Requirements on Deliberation (chapter 2)

Levi's epistemic requirements on deliberation are put forward in the following. I will revise and defend these in developing my epistemic model of free deliberation.

Levi, Isaac. 1986. *Hard choices*. Cambridge: Cambridge University Press.

———. 1997. *The Covenant of Reason*. Cambridge: Cambridge University Press.

———. 2000. Review of James M. Joyce: The Foundations of Causal Decision Theory. *Journal of Philosophy* 97(7): 387–402.

Similar requirements are also explored and then defended in the following:

Gilboa, I. 1999. Can Free Choice be Known. in C. Bicchieri, R. Jeffrey and B. Skyrms (eds.) *The Logic of Strategy*. Oxford: Oxford Uni. Press.

Ginet, Carl. 1962. Can The Will be Caused? *The Philosophical Review* 71(1): 49–55.

Jeffrey, Richard C. 1977. A Note on the Kinematics of Preference. *Erkenntnis* 11(1): 135–141.

Goldman, Alvin. 1970. *A Theory of Human Action*. Englewood Cliffs, NJ: Prentice-Hall.

Holton, Richard. 2006. The Act of Choice. *Philosophers' Imprint* 6(3): 1–15.

Schick, Frederic. 1979. Self-knowledge, Uncertainty, and Choice *British Journal for the Philosophy of Science* 30: 235–252.

Shackle, G. L. S. 1958. *Time in Economics*. Amsterdam : North Holland Pub. Co.

Spohn, Wolfgang. 1977. Where Luce and Krantz do Really Generalize Savage's Decision Model. *Erkenntnis* 11: 113–134.

Nelkin, Dana. K. 2004. Deliberative Alternatives. *Philosophical Topics*. 32(1&2): 215–240.

3. Competing Accounts of Free Deliberation (chapter 3)

The following two works argue against the epistemic requirements on deliberation I defend, and posit competing accounts of free choice in deliberation.

Joyce, James M. 2002. Levi on Causal Decision Theory and the Possibility of Predicting One's Own Actions. *Philosophical Studies* 110(1): 69–102.

———. 2007. “Are Newcomb Problems Really Decisions?.” *Synthese* 156: 537–62.

Rabinowicz, Wlodek. 2002. Does Practical Deliberation Crowd Out Self-Prediction? *Erkenntnis* 57(1): 91–122.

Joyce supports his account by appealing to Velleman's account, presented in the following:

Velleman, J. David. 1989a. *Practical Reflection*. Princeton: Princeton University Press.

———. 1989b. Epistemic Freedom. *Pacific Philosophical Quarterly* 70: 73–97.

———. 1992. What Happens When Someone Acts? *Mind* 101(403): 461–481.

———. 2000. *The Possibility of Practical Reason*. Oxford: Oxford Uni. Press.

4. Reductivist Accounts of Causation (chapters 1 and 4)

The classic expression of scepticism for treating causation as fundamental is:

Russell, Bertrand. 1912–13. On the Notion of Cause. *Proceedings of the Aristotelian Society*, New Series, 13: 1-26.

A general useful collection on various reductionist accounts of causation is:

Price, H. and Corry, R. (eds.) 2007. *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*. Oxford: Oxford University Press.

The reductivist accounts of causation I will consider are David Lewis' and James Woodward's. (I may consider other manipulationist accounts as well, such as Pearl's.) Lewis offers a reductivist account of counterfactuals as well, while Woodward, takes counterfactuals as primitives.

Lewis, David. 1973a. *Counterfactuals*. Oxford: Blackwell.

———. 1973b. Causation *Journal of Philosophy*, 70: 556–67. (and postscripts in (1986a): *Philosophical Papers: Volume II*. Oxford: Oxford University Press.)

———. 1979. Counterfactual Dependence and Time's Arrow. *Nous* 13: 455–76.

Woodward, James. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

———. 2007. Causation with a Human Face” in Price, H. and Corry, R. (eds.) *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*. Oxford: Oxford University Press: 66–105.

Jenann Ismael and Huw Price both explicitly explain aspects of causation by reference to decision contexts. Aspects of their individual accounts are presented in the following:

Ismael, Jenann. 2007. Freedom, Compulsion and Causation. *Psyche* 13(1): 1–11.

———. 2012. Decision and the Open Future. in A. Bardon (ed.) *The Future of the Philosophy of Time*. London: Routledge.

Price, Huw. 2012. Causation, Chance and the Rational Significance of Supernatural Evidence. *Philosophical Review* 121 (4):483-538

———. 1991. Agency and Probabilistic Causality. *British Journal for the Philosophy of Science* 42: 157–76.

Menzies, Peter and Price, Huw. 1993. Causation as a Secondary Quality. *British Journal for the Philosophy of Science* 44: 187–203.

5. Causal Asymmetry and Statistical Mechanics (chapter 5)

Objections to Lewis' account based in statistical mechanics:

Andy Egan. 2007. Some Counterexamples to Causal Decision Theory. *Philosophical Review* 116 (1): 93-114.

Accounts that rely on the Past Hypothesis and statistical mechanics to explain an asymmetry in causation include:

Albert, David Z. 2000. *Time and Chance*. Cambridge, Mass.: Harvard Uni. Press.

———. Forthcoming. The Sharpness of the Distinction between Past and Future in A. Wilson (ed.) *Asymmetries of Chance and Time*. Oxford: Oxford Uni. Press.

———. Unpublished. *Life after Physics*.

Carroll, Sean. 2010. *From Eternity to Here*. New York: Dutton.

Kutach, Douglas. 2002. The Entropy Theory of Counterfactuals. *Philosophy of Science* 69(1): 82–104.

Loewer, Barry. 2007. Counterfactuals and the Second Law. in H. Price & R. Corry (eds.) *Causation, Physics, and the Constitution of Reality*. Oxford: Oxford Uni. Press.

Price, Huw. 2002. Boltzmann's Time Bomb. *British Journal for the Philosophy of Science* 53: 83–119.