Time, Flies, and Why We Can't Control the Past

Alison Fernandes

David Albert explains why we can typically influence the future but not the past by appealing to an initial low-entropy state of the universe. And he argues that in the rare cases where we can influence the past, we cannot use this influence to knowingly gain future rewards: so it does not constitute control. I will introduce an important new case in which Albert's account implies we can not only influence the past but control it: a case where our actions in the present are reliably correlated with several events in the present and past. To deal with such cases, we need to appeal to epistemic conditions on deliberation: being agents requires our decisions being epistemically undetermined at the time we make them. In a world with the past-hypothesis, this implies that deliberation will typically come prior to decision. Once deliberation in this direction is established, correlations towards the past cannot then be exploited for control. To explain why we cannot effectively control the past, we need to appeal to deliberation, whether as part of a defence of Albert's account, or used independently to explain the asymmetry of control.

1. Introduction

David Albert explains why we can typically influence the future but not the past by appealing to an initial low-entropy state of the universe, a postulate he calls the 'Past Hypothesis' (2000). And he argues (forthcoming), that in the rare cases where we can influence the past, we cannot use this influence effectively, to knowingly gain future rewards. So it does not constitute control. But there is an important case Albert fails to consider: a case where our action in the present is reliably correlated with several events in the past, and can knowingly

be used to gain future rewards.[1] To deal with this case, we need to appeal to epistemic conditions on deliberation: being agents requires our decisions being epistemically undetermined at the time we make them. In a world with the Past Hypothesis, this implies that deliberation will typically come prior to decision. Once deliberation in this direction is established, correlations towards the past cannot then be exploited for control. Deliberation is required to explain why we cannot effectively control the past, whether as part of a defence of Albert's account, or used independently to explain the asymmetry of control.

I begin with Albert's explanation of the asymmetry of influence—why the future depends on what we do now in a way the past does not (§2). I then argue against his treatment of exceptional cases. His response to a case presented by Mathias Frisch fails, because it introduces an unexplained asymmetry of rewards (§3). Moreover, he cannot deal with a new importantly different case by appealing to conditions on rewards (§4–5). In this new case, it seems we *can* knowingly influence the past to gain future rewards. To deal with this case, we need to appeal to evidential undermining and epistemic conditions on deliberation that explain why agents deliberate before they decide (§6). By appealing to deliberation, we can once again explain the asymmetry of control.

If we can recover Albert's explanation, we revive his more general project of unifying and explaining a range of asymmetric phenomena. This follows a tradition, beginning with Boltzmann's work in of statistical mechanics, of explaining apparently fundamental phenomena in scientific terms: why ice cubes in glasses of water tend to melt (a

[1] My target here is Albert, but related problems arise for the entropy-based accounts of Loewer (2007, 2012) and Kutach (2002). For further criticism of Albert's account, see Frisch (2007, 2010) and Price and Weslake (2009).

thermodynamic asymmetry), why our knowledge of the past is so different from our knowledge of the future (an epistemic asymmetry), why we can influence the past but not the future, and why causes come before their effects. What typifies this program is attempting to account for these asymmetries by appealing merely to time-symmetric dynamical laws, assumptions and methods, and the Past Hypothesis.[2] The program uses a form of explanation that is already well-accepted in scientific explanations of the second law of thermodynamics. And by tracing a range of asymmetric phenomena back to a single boundary condition, we can understand at a deep level how these phenomena relate. So this kind of explanation is worth preserving.

2. The Asymmetry of Influence

Albert's explanation of the asymmetry of control proceeds in two parts.[3] In the first, Albert appeals to the Past Hypothesis to explain why the present contains many more means of influencing ('causal handles') on macroscopic aspects of the future than the past. In the second part (considered in the next section), Albert argues that any causal handles there are on the past cannot be used effectively—we cannot know about them or use them to gain rewards. So even if we can influence the past, we can't control it.

The first thing to note is that talk of causal handles as 'ways of influencing the future' should be compatible with the deterministic character of the dynamical laws. Here I take the laws to be Newtonian laws of motion. Although such laws are not exceptionless or fundamental,

[2] For an overview of foundational work in statistical mechanics, see Sklar (1993). For work broadly in this program, see Reichenbach (1956), Horwich (1987), Price (2002), Callendar (2004), Greene (2004), Penrose (2005), Loewer (2007, 2012), Carroll (2010) and references therein, and for recent criticism, Winsberg (2004), Earman (2006), Price and Weslake (2009) and Frisch (2007, 2010).

[3] Albert sometimes refers to a merely *apparent* asymmetry. I take it the asymmetry is real, for any reasonable concept of control, but this choice does not significantly affect the explanation.

whatever laws replace them will have to recover Newtonian laws on everyday length and energy scales, leaving the explanation largely intact. When we claim that manipulating a local feature of the present is a way of influencing the future, we are not claiming, contrary to fact, that two different futures are compatible with the present state and the laws. What we are claiming is that, counterfactually, *if* the present were otherwise than it is, *then* the future would be different. This in turn means we need some way of evaluating counterfactuals.

Albert proposes the following method.[4] Start with phase space, a continuous space containing six dimensions for every particle in the system, one for each dimension of position and momentum. Points in phase space represent possible microstates of the system at a time. Volumes of phase space represent possible macrostates. Given deterministic laws, some of these microstates are compatible with the Past Hypothesis: with the universe beginning in 'whatever particular low-entropy highly condensed big-bang sort of macrocondition it is that the normal inferential procedures of cosmology will eventually present to us' (2000, 96). The Past Hypothesis claims the universe began in a particular low-entropy macrostate.[5] To evaluate a counterfactual, take the location of the actual world in phase space at the present time, and consider the closest world (as measured in phase space at that time) that satisfies the following:

1) The antecedent of the conditional is satisfied,
2) The Past Hypothesis is true,

[4] Albert claims that any reasonable time-symmetric method of evaluating counterfactuals yields the same asymmetry (2000, 125). Loewer's account (2007) is broadly compatible. Lewis' account (1979) is symmetric, but beset by further problems (see Elga 2001).
[5] We will need to amend this claim if the concept of entropy does not apply at the start of the universe (Albert 2000, 85 and Earman 2006, 412).

3) The world's macrohistory, given its macrostate and the Past Hypotheses, is assigned a reasonable probability by the statistical postulate.

The statistical postulate in the third condition takes probabilities of volumes of phase-space at a time to be uniform over the standard Lebesgue measure. This implies, roughly, that for any two volumes of the same size, the system is as likely to be in one as the other. The third condition excludes cases where the closest world is an anomalous one—in an improbable microstate, i.e. with an unusual macrohistory, given its current macrostate and the Past Hypothesis. We will see some of the interesting results of this condition below. Once the world has been selected, we evolve it in time according to the dynamical laws, and see whether the consequent is satisfied. The counterfactual is true if and only if the consequent is satisfied.

Note that any change to the present microcondition of a world will necessarily change its past as well as future microhistory. Any counterfactual change to the present is necessarily a change to the past and future. This means that if we are to succeed in explaining the asymmetry of influence, we need to restrict the changes to the past, present and future being considered, by restricting the possible antecedents and consequents of the counterfactuals. Albert restricts the consequents to localised areas of the world that can be characterised macroscopically in relatively simple everyday language (2000, 121). He restricts the antecedents by introducing what he calls a 'fiction of agency'. This is a 'primitive and un-argued-for' conception of what lies under our 'direct and unproblematical and unmediated control' (2000, 128). While 'fiction' suggests such control is illusory, it need not be—for the moment we can simply take it as a primitive conception of what we take to be directly under

our control. This 'black-box' conception can be filled out in various ways—to allow us direct control of our limbs, for example, or our tendons or the electrical nerve impulses in our brains. Albert claims that under any reasonable conception, direct control will be localised to a very small area of the universe.

Given these restrictions, counterfactuals determine what aspects of the present, past and future we influence. Albert claims that if the antecedents are features that are under our direct control and the consequents are readily characterisable macroscopic features, the present contains many more 'causal handles' on the future than the past. The turning motion of my hands on the steering wheel, for example, is a causal handle on my car beginning to turn left in the next moment. But it is not a causal handle on the car's previous motion. This reasoning is based on the claim that the Past Hypothesis accounts for the asymmetry of *records*: it explains why we can reliably read local macroscopic features of the present and our memories as reliable indicators of the macroscopic past, but not the future (2000, Chapter 6). Assuming this, and that the same inferential procedure are used for reading records and evaluating counterfactuals, we get the asymmetry of influence. The method for evaluating counterfactuals requires leaving most of the actual present as it is, including records. If these records are reliable, when we evolve the system backwards in time, the past inferred to will be very much like the actual one. But in regards to the future there is no such guarantee. If the present contains no records of the future, a small change in one part of the present may well bring about large changes in the future—the present state of the world does not constrain these changes to the same degree.

Consider an example. In the actual world, my cousin saws away at a tree until the tree is almost ready to topple, and then waits for my signal. I signal by waving my arm, she makes the final cut and the tree comes crashing down. If I had not waved my arm, could I have influenced anything significant about the future? When we consider the nearest counterfactual world in which the antecedent is satisfied, where I do not wave my arm, and evolve this forward, it is likely the tree does not come crashing down. No signal is given, and my cousin does not continue sawing. But could I have influenced anything significant about the past? The nearest counterfactual world in which the antecedent is satisfied in the present has me not waving my arm. But it is likely to have lots of *records* of the sawing having begun—my memories, wood chips lying on the ground, birds flying startled in the air. If evaluating counterfactuals involves the same inferential procedures as reading records, these reliable records keep the past as it was. Given the restriction imposed by the Past Hypothesis, it is likely that my cousin has still been sawing away at the tree.

This same type of reasoning works for any macroscopic event in the past for which there are records in the present. So it seems that if there are records in the present that are not under our direct control, we cannot influence these aspects of the past, and so our influence of the past is severely restricted in a way our influence of the future is not. By explaining the asymmetry of records, the Past Hypothesis thus explains the asymmetry of influence.

3. Dealing with Exceptional Cases

However, cases presented by Adam Elga (2001), Douglas Kutach (2002) and Mathias Frisch (2010) suggest that Albert's account entails that we *can* influence the past. So Albert can't explain the asymmetry of influence. Albert responds by moving to consider *control* of the

past, rather than *influence*, and introducing further conditions. While some are problematic, conditions on control are enough to deal with these exceptional cases—but not with a further new case I present.

Elga and Kutach introduce the following case.[6] There is a large-scale event in the past, such as the sinking of Atlantis, for which there are no records in the present. This means that small changes in the present, such as the movement of my finger, may be correlated whether Atlantis exists. Given we control such small changes in the present, Albert's account implies we control past macroscopic events.

Albert (forthcoming) and Loewer (2012, 128) respond by arguing that while there may be counterfactual dependence of the past even on the present, this does not constitute *control*. This is because there are no macroscopic correlations we could become aware of between what we control in the present and the past macrostate. If there were, there would be records in the present, and this case relies on the absence of records.

This response succeeds, but only by conceding that an asymmetry of counterfactual dependence is not enough to explain the asymmetry of control. Instead, we need further epistemic conditions on what counts as effective control: it is control we can know about. But such epistemic conditions are plausible. We don't have control of an event merely because it is correlated with our action. For example, I do not control what a detective records merely because he has decided to write down my shopping choices (to take a case from Anscombe). I need to at least know about this correlation in order to effectively use it

[6] The case is based on observations in Elga (2001), cited in Loewer (2012) and Albert (forthcoming).

to bring about my ends, and so for it to constitute control. With this added condition, Albert's account survives Elga and Kutach's challenge.

Frisch (2010) raises a separate challenge to Albert's account. He introduces a case where the unique record of an event is under an agent's direct control. Say I am playing a piano piece, and come to a point in the music where I can decide whether to play the first or second ending—the decision being under my direct control. Stipulate that I have no conscious memory of what music I have played and there are no external records present. But I am a reliable pianist. Therefore, whatever decision I make counts as a record of where I am up to and what music I have played. So by having control of my decision in the present and my decision being a unique record, I have control over what music I have played in the past. So we have a counterexample, and Albert's account does not deliver the asymmetry of influence.

Albert (forthcoming) responds to Frisch's case by adding another condition on effective control. He argues that even if the agent knows he can influence the past, he cannot use this influence for any further benefit, beyond what he secures through influencing the present. If the pianist wants to finish his playing sooner, for example, he could influence the past so that he began the piece earlier. But he could just as easily decide to play the second ending and finish playing. The past event only influences the future via a feature the agent already has direct control over. It gives the agent no additional benefits. So this is not a case of effective control, even if it's a case of influence.

But this response fails. Albert only considers additional benefits in the future. This introduces an unexplained asymmetry. If we consider benefits in the past as well, influencing the past in the Frisch case may give the agent additional rewards. For example, he may have spent less time bored at the piano, if he started his playing earlier. Influencing the past is essential for securing this gain. For Albert's response to work, this asymmetry in what benefits we consider must be explained in a non-circular fashion. And even if this can be done, there is an important new case for which this response is ineffective.

Before we move on to the new case, we should note that there is a better response to Frisch available. To effectively control the past, it is not enough to know that in some cases, my decision is correlated with a past event. To know I control the past, I must have some way of checking whether I have controlled it, on occasions when I *attempt* to do so. I cannot do this in the Frisch case. There is no time at which I know both my decision and my past playing by independent means. Because my decision in the present is the only record of the event, I cannot determine whether I have controlled the past. And so this is not control I can know about. So Frisch's attacks fails. Albert's explanation of the asymmetry of control survives. But there is a new case that evades even this response.

4. The Fly Case

The Elga-Kutach and Frisch cases can be dealt with. But there is another case that threatens Albert's explanation of the asymmetry of control. Here it is: in the actual world, a fly flies in front of my face at t_1, and at t_2 I swat it away. Then consider the counterfactual: if I had not swatted at t_2, the fly would have been somewhere else at t_1. If this counterfactual is true, and I can knowingly exploit this counterfactual dependence, I have control of the past. And so

Albert's account does not deliver the asymmetry of control. In this section I will argue the counterfactual is true, and in the next, that I can knowingly exploit this counterfactual dependence.

To evaluate the counterfactual, we consider nearby worlds in which I do not swat, the Past Hypothesis is true, and the macrohistory of the world is statistically normal. Perhaps I don't swat because a stray thought or itch distracts me. But say I am not the type to get distracted by stray thoughts or itches—since I was a young child, many varied experiences have led me to treat flies as small flying vermin, to be efficiently swatted away whenever they appear. Moreover, there are lots of records of these harrowing experiences, and my reliability at swatting. Given that these records are reliable and kept intact, my swatting will itself count as a record of the fly's location in the past. If I hadn't swatted at t_2 it would be because the fly was out of range or sight at t_1. Our counterfactual looks true, set to give me influence on the past.

The counterfactual would be false if the present contained many large-scale records of the fly's actual position at t_1. Keeping these records intact and reliable would prevent the location in the past from changing—as we saw in the tree-sawing case above. But no entomologists are videoing the fly's activities. The fly's location at t_1 is a small event, perhaps only recorded in its location at t_2, the location of some air molecules and my memory. These records are small enough to be changed in a nearby world where I don't swat. And these changes are required if my tendency to swat and these records are to remain reliable.

This is the important difference between the fly case and the Frisch case. In the Frisch case, there are no macroscopic changes beyond those explicitly specified in the antecedent. The agent's influence in the present is limited to what he controls directly. Not so in the fly case. By directly controlling a record of the fly's location, the agent influences other records in the present. And this allows him more wide-ranging control of the past and future, as we will see in the next section.

What is doing the work in making the counterfactual true is the third condition: that the world's macrohistory, given its macrostate and the Past Hypothesis, is assigned a reasonable probability by the statistical postulate. There are worlds in which there are lots of records of my reliability as a flyswatter and yet my swatting behaviour is uncorrelated with the location of the fly in the past. But these are all unusual worlds, assigned a low probability by the statistical postulate, not compared to each other, but compared to a given standard. We should not infer to such worlds. Similarly there are worlds in which the fly's location at t_1 changes, but not its location at t_2: the fly flies incredibly fast. But again, these are unusual worlds. We should not infer to them.

There is a nearby world in which all these records remain reliable. It is a world in which I do not swat and the fly's location at t_1 and at t_2 change. This is the world I should infer to. These changes are required to satisfy the antecedent, and make the world's history sufficiently probable, given its macrostate. My swatting remains a reliable record of the fly's location. And so by directly controlling my swatting, a record of the fly's location, I influence the fly's location in the present and past. Albert's account does not yet deliver the asymmetry of control. I will now defend this counterexample from a number of preliminary objections.

Objection 1: The relevant counterfactual is false. The third condition only requires us to pick a microstate that is probable *with respect to other microstates in the same macrostate*. Given Frisch's interpretation (2007, 2010), no macrostates are ruled out by this condition. In the nearest world, nothing macroscopic outside the antecedent changes. So the fly's location in the past does not change.

Response: Albert's account is consistent with this reading of the third condition, but it favours my own. My reading is more in keeping with Albert's appeal to 'normal procedures of inference' and his defence of these three conditions as capturing how we ordinarily infer (2001, 130). It is normal to infer from a record of one event to other records of the same event at that time. My morning newspaper allows me to infer the news of yesterday as well as the contents of your newspaper today. If counterfactuals are to capture such inferences, those like the following should come out true: if my paper had been different, yours would have been. If this counterfactual is to come out true, we can't rule out making changes in the macroscopic state outside the antecedent. In addition, Albert (forthcoming) more explicitly requires us to: 'Find the possible world which is closest to the actual one, as measured by distance in phase-space, at the time of the antecedent…'. If there are to be no macroscopic changes outside the antecedent, this requirement is difficult to make sense of. Why should closeness within the same macrostate matter? The condition makes better sense if its work is to *minimise* such macroscopic changes, not rule them out altogether.

Objection 2: Albert's intentions aside, we should follow Loewer (2007) and Kutach (2002) and not allow macroscopic changes outside the antecedent when evaluating counterfactuals. So the counterfactual is false.

Response: Satisfying antecedents nearly always involve making changes elsewhere in the system, under any method of evaluating counterfactuals. For example, if my hand is to be located in region *x* instead of *y*, air molecules will have to be moved out of *x* and into *y*. This is not an unusual feature of Albert's method. Moreover, the method gives us a useful guide to what other changes are required to satisfy the antecedent. Because worlds with vacuum pockets around me are very unusual, we should not infer that were my hand at *x*, there would be a vacuum pocket at *y*.

In addition, allowing the macrostate outside the antecedent to change is in keeping with our normal inferential reasoning. We reason from states of one system to states of another system at the very same time. If the counterfactual method is to preserve this reasoning, it must allow for changes in the antecedent to be correlated with macroscopic changes outside the antecedent. While ultimately we may want a method that doesn't allow for changes outside the antecedent, the more informative project is to consider how we get to such a method without first simply stipulating against such changes. Price and Weslake make a similar point (2009, 426).

Objection 3: Condition 3 can only rule out exceedingly unlikely worlds. A world where I swat and the fly does not move is not sufficiently unlikely.

Response: If condition 3 is to ensure good inferential reasoning, it should not be too weak. We should not infer that measuring devices always break down when we consider changes to their records. While unusual events do happen in our world, and reliable devices do break down, we should assume they always do, if there are other nearby worlds available, in which

recording devices remain reliable. Otherwise the method for evaluating counterfactuals will often lead to unhelpful results.

Objection 4: If we're requiring records remaining reliable, why should the size of the record matter? Why not also consider *large* macroscopic changes outside the antecedent?

Response: While such a method is appealing, my target is Albert. Albert follows Lewis (1979) in thinking that changes from the actual world should be minimised. What's important is that even with this requirement, counterexamples can be generated.

Objection 5: The fly's location is a small event. The influence on the past is only minimal, and so Albert's explanation succeeds for the most part.

Response: While the fly's location is a small event, a small change in the near past can be parlayed into a much larger change in the more distant past. If the small event I control is a unique record of a larger more distant event, say the last remaining newspaper record of a 19th century fire, I control the larger event as well. Nothing in the framework itself rules out such large-scale changes—or us training ourselves and others to develop responses that can be exploited in this way.

Objection 6: The setup for the case itself is too fragile—add some large records in the present, and the influence on the past is destroyed. These cases are not sufficiently common and robust to challenge Albert's explanation.

Response: While the case is fragile, this is not enough to dismiss it. We should consider why we can't make our influence on the past more robust, as we can our influence on the future. The response I recommend below can account for this fact.

Objection 7: Agents aren't reliable responders. Their neurological makeup is fragile and easily changes in counterfactual worlds. Their decisions and actions aren't records, and so don't allow for influence of the past under Albert's set up.

Response: It is essential for the fly case that records of the agent's reliability are recorded in large-scale macrofeatures of the world, such as in eyewitness reports and my four-volume treatise detailing my dedication to fly swatting. These macrofeatures don't easily change and are reliable evidence for the behaviour and neurology of agents.

Objection 8: Agents aren't reliable responders unless they observe the events they respond to. Having observed them, they should hold them fixed in evaluating counterfactuals.

Response: My swatting may be an unconscious, instinctive action and yet reliable. But more importantly, we should ask why this is true of our reliable correlations to the past, but not the future—why we don't need to observe a future states in order for them to be correlated with our decisions now. Calling correlations with the past 'responses' hides this asymmetry. Once this asymmetry is explained, the response suggested is compatible with my own (§6).

5. Can the Fly Case Be Used for Effective Control?

Having considered some preliminary objections, how else might we respond to the fly case? Albert responded to the Elga-Kutach and Frisch by invoking conditions on effective control. While these cases involve counterfactual dependence and influencing the past, I cannot knowingly use this influence to gain rewards. So they do not threaten an asymmetry of control. But I will show how it seems we can knowingly gain rewards in the fly case, and it remains a threat to Albert's account.

To show this, we can enrich the fly case by adding a sensor that records the location of the fly at t_1 and displays this location on a screen at t_2. By stipulation, the screen is the only record of the fly's location from the sensor that remains at t_2—the rest of the device resets. Stipulate this record is small in size and changes in the counterfactual world. This means both the record and my hand swatting are reliably correlated with the presence of the fly in an area at t_1. Say I am offered a reward (anytime before t_2) for having this area free of flies at t_1. I will attempt this at t_2, the sensor will tell whether I have succeeded and I will be rewarded when the screen is checked after t_2. From what has been said so far, I can influence the fly's location to gain the reward. So I can control the past, and Albert's account doesn't yield an asymmetry of control.

My response to the Frisch case will not work here. In the Frisch case, there is no independent means of checking whether I have controlled the past. The only record of the past event is a record I control directly, and there is no way to confirm that this record remains reliable. What is importantly different about the fly case is that there are records of the consequent *outside* the direct control of the agent—the fly's location in the present, and

the record on the screen. This means there is an independent means of checking whether I have controlled the past.

Albert's response to the Frisch case will not work here either. His objection to the Frisch case was that any rewards the agent gains in the future are due only to his direct control of the present, not his influence on the past. Not so in the fly case. Here my mere ability not to swat, uncorrelated with the absence of the fly, is not enough to secure the reward. I have to influence the fly's location in the past. Otherwise the sensor's record in the present will not be changed, and I will not receive the reward. Again, this is because there are records of the consequent outside the direct control of the agent. The agent can influence these records, but only by influencing the past.

Nor is Albert's response to the Elga-Kutach case effective here. In the Elga-Kutach case, there are no records of the past event in the present, and so no macroscopic correlations between my action and the event I could become aware of. But in the fly case, there are such records. There are macroscopic correlations between my swatting now and the fly's location in the past, correlations I can know. So it seems I *can* control the past to gain the reward.

But there is a more promising response: I cannot use my influence of the fly's location for gain, because any attempt to exploit it destroys the correlation on which it is based. If I commit myself to gaining the reward and therefore decide not to swat my hand, the stillness of my hand is no longer a reliable record of the fly's absence. Instead, it is a reliable record of my desire for or reason to gain the reward. In the nearest counterfactual world, the macrohistory will then not be one in which the fly is somewhere else, but merely one in

which I wish or have reason to make this so. More generally, if the antecedent is brought about in order to bring about the consequent, the antecedent is no longer a reliable record of the consequent, but only of a desire or reason to bring it about. So the counterfactual is false and I do not control the past. I call this 'evidential undermining'.

Evidential undermining works equally well against the Frisch case. If I decide to play the second ending in order that I started playing earlier, and so will finish sooner, my playing the second ending is no longer a reliable record of my having begun earlier—just of my desire that this is so. While Frisch is right to claim that we can learn inductively of correlations between our decisions and past events (2010, 31), this is not the same as learning of correlations we can make use of; these need to be correlations which are robust when we attempt to use them to control the past.

Note that this evidential undermining is analogous to responses made by evidential decision theorists to medical Newcomb cases (Price 1986, 1991). The general idea is that the evidential correlations an agent makes use of in decision must survive her knowledge of her deliberation—a deliberation that includes a belief in the correlation. In medical Newcomb cases, the agent has no reason to believe that correlations between her actions now and past events will be preserved if she decides to act on the basis of a belief in this correlation (in contrast to traditional Newcomb cases). So evidential decision theory gives her no reason to decide on past events. The corresponding response in the fly case is that evidence for a correlation between the fly's presence and my swatting, is not evidence for a correlation between a fly's presence and my swatting in cases where I decide not to swat in order to exploit this correlation.

This response is largely correct—here I agree with Price and Weslake (2009). But it is inadequate in its current form. It relies on an asymmetry—that agents in our world deliberate *before* they decide. It is because deliberative reasoning comes temporally prior to decision that it can disrupt correlations between decision and past states, but not correlations between decision and future states. This may be due to general features of influence: that influence propagates through time. Or it may be due to features peculiar to deliberation. But it is straightforwardly true. If an agent's deliberative structure were reversed, such that she acted, then decided, then deliberated, correlations between her decision and past states would be preserved under counterfactual changes.

So if evidential undermining is to succeed as a response, we must explain why deliberation comes before decision. In the final section, I will explain this fact, by appealing to epistemic conditions on deliberation, and an epistemic asymmetry. What will emerge is that while we can appeal to the above response to defend Albert's account of the asymmetry of control against cases like the fly case, another option is available. We can explain the asymmetry of control more directly by the fact that agents deliberate before they decide and their deliberation undermines correlations in the same temporal direction. Moreover this explanation requires a more limited version of the epistemic asymmetry than Albert's own.

6. Deliberative Agency

I will now explain why deliberation comes temporally prior to decision. I will argue for two epistemic conditions on deliberation (*ignorance* and *efficacy*) that allow us to explain a key feature of deliberation: its seriality. *Seriality*, combined with an epistemic asymmetry, explains this deliberative asymmetry and so the asymmetry of control.

Ignorance

I take it to be a condition of deliberation that an agent cannot be certain of her decision at the time she is deliberating and as she decides.[7] If an agent is deliberating whether to *A*, she cannot be certain that her deliberation issues in her deciding to *A*: decisions are *epistemically undetermined*. Why think that decisions satisfy this requirement? While many theorists simply note the condition as necessary, a promising response appeals to the function of deliberation: there is no point deliberating if an agent is already certain of her decision. Holton, for example, argues that choice scenarios arise in cases where we are unsure of what to do—if we were sure, there would be no need for the decision (2006). Ginet makes similar points (1962).

Accepting this requirement places an *ignorance condition* on decision. There cannot be connections between and agent's decisions and states of the world (including her beliefs and desires), which allow her to infer her decision while deliberating and deciding. In particular, if she is certain of other states of the world, she can't also be certain of the correlations between these states and her decision. *Ignorance* can be satisfied if the correlations between states of the world and decisions are complicated and difficult to track. This will be the case if our internal make up is complex, constantly changing and hidden from view. Think of a fireman, and whether he can take the heat of the floor as evidence he will decide to leave a burning building (Holton 2006). In order to make an accurate inference, he must take into account the nature of his values and desires, the way he balances the risks involved against

[7] Defenders of this condition include Hampshire and Hart (1958), Ginet (1962), Taylor (1964), Schick (1979), Levi (1986), Holton (2006) and Ismael (2007, 2012). The general shape of this account of agency draws much from Ismael's work.

the possibility of saving the building, and the significance of all this evidence. In addition, his inference must take into account what effect the inference will have on his decision—making a prediction can change his evidence about what he will decide to do. Because of this, his decision is very unlikely to be susceptible to simple modelling. *Ignorance* can be satisfied by these kind of epistemic limitations.

Efficacy

There is a second epistemic condition on decision: agents must take their decisions to be *efficacious* in producing their results. An agent, while deliberating whether to *A*, must believe that her decision to *A* will result in her *A*-ing: her decision must be good evidence for her A-ing. This does not imply a causal connection between the decision and result. It is a minimal epistemic condition only, which many defend (see footnote 7).

These two conditions explain how features of the world, particularly bodily movements, appear to be under our direct control. So we can do better than Albert, and not take direct control as a primitive. As far as our decisions are epistemically undetermined, other states of the world we take to be reliably correlated with them will appear under our direct control. Further conditions may be needed to ensure direct control. But as necessary conditions, *ignorance* and *efficacy* partly determine the appropriate scope of our conception of direct control: it will include only features that we take to be reliably correlated with our decisions.

Seriality

These two conditions on decision explain a key feature of deliberation: its *seriality*. This is its ordering in terms of deliberation, decision, action, as schematized in figure 1 below. Note

that this is not yet an asymmetric temporal ordering—we're not presuming that decision falls *before* action in time. Decision might come after. But *seriality* does imply that decision falls temporally between action and deliberation.

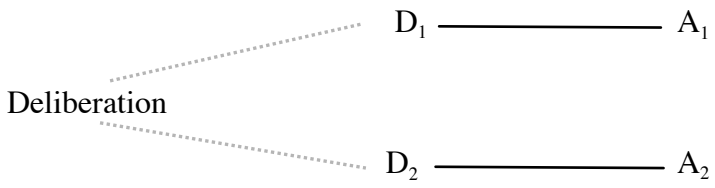

Figure 1: The seriality of agency. While deliberating, the agent is uncertain of her decisions (shown by a dotted line). Decisions are epistemic possibilities (D_1 or D_2). But she takes there to be a correlation between her decisions and her actions (A_1 and A_2), represented by a solid line.

Here is how *seriality* results. *Ignorance* implies that an agent cannot be certain of her decision while deliberating and deciding. *Efficacy* implies she must be certain of the correlation between her decision and her action while deliberating. Together they imply than an agent cannot be certain of her action while deliberating—given *efficacy*, this would violate *ignorance*. But to satisfy *efficacy*, the agent also needs some way of becoming aware of correlations between her decisions and actions in general. And this means there must be a time at which she knows both her decision and her action. Unless an agent typically becomes ignorant of her decision between acting and deciding (a very unusual epistemic structure indeed), she cannot typically deliberate at times between deciding and acting—this would violate *ignorance*. And she cannot act between deciding and deliberating—this would prevent there being a

time at which both decision and action are known. So agency must take this serial form. It is a form that allows there to be a time at which both decision and action are known, but which does not prevent the agent deliberating at another time.

It does not yet follow that agency is oriented in a particular temporal direction—only that decision comes between deliberation and action. However, given the epistemic asymmetry Albert explains (or even a more limited epistemic asymmetry), *seriality* implies agency will typically lie in a temporal direction that runs past–future. *Ignorance* and *efficacy* can be satisfied in a forward temporal direction. Our decision-making is sufficiently complex to satisfy *ignorance*, as we saw in the fireman case. And records in the form of memories allow agents to pick up on general correlations between their decisions and actions, allowing *efficacy* to be satisfied. Not so in the backward temporal direction. Here, the reliability of records prevents *ignorance* being satisfied. If our macroscopic acts occur in the recent past, we will typically have records of them in the present in the form of memories. Given *ignorance* this prevents us deliberating about these acts: the decisions are not epistemically undetermined. An agent could still take the mental state to be epistemically undetermined, by ignoring or forgetting the correlation with action. But then it would not satisfy *efficacy* and would longer count as a decision. And note that if the temporal distance between the decision and the action becomes too great, the reliability of their connection is likely to be lost and so *efficacy* will not be satisfied.

It is because there are readily available records of macroscopic events in the recent past, but not the future, that deliberation is typically directed towards the future and we deliberate before we decide. Records of the past prevent decision being directed towards the past. But

they don't prevent decision from being directed towards the future, and they allow agents to track correlations between decision and action. This argument does require an epistemic asymmetry. But we have independent grounds to accept such an asymmetry, whether it is a general asymmetry of records (Albert 2000) or something limited to records like memories. However, if a suitable epistemic asymmetry can be explained in terms of the Past Hypothesis, as Albert attempts, we also have an argument for why the Past Hypothesis is relevant to the direction of deliberation. It is relevant because it explains features of our epistemic access to the world (*contra* Price and Weslake 2009).

The seriality of agency and an asymmetry of records explain why agents deliberate before they decide and why their decisions are typically directed towards the future. But we haven't yet explained why agents *never* deliberate about the past. There are two ways that agents could generalise from the typical case and adopt a conception of agency that rules out deciding on past states of affairs. One is by overgeneralising and believing there is no counterfactual dependence between decision and macroscopic states they could have access to at the time of deliberation. Given all the past seems "in principle" accessible, we cannot influence the past. So-called 'bilking' arguments against influence of the past exploit this idea (Dummett 1964). A second is by overgeneralising and believing there is no counterfactual dependence of past macroscopic states on present decisions (mentioned Frisch 2007, 365). Both these conceptions of agency can explain what seems odd about the fly and Frisch cases and why we take influence of the past to be *impossible*.

But we need not go so far. We can combine the deliberative asymmetry with our previous response to the fly case to finally explain the asymmetry of control. In §5, I argued that

agents who deliberate prior to decision will be unable to exploit correlations between their decisions and states further in the past. If they attempt to use such correlations to control the past, they will undermine them. Their decisions will only be evidence for their desires or reasons, rather than for the past states they hope to control. So as far as an agent's evidential map encodes correlations she can use, no correlations between decisions and previous states will appear. While she can believe there are evidential correlations between her decision and previous states, they are not ones she can exploit to achieve her ends.

So agents cannot freely establish correlations between their decision and past events and use these to control the past. In the fly case, for example, if I decide to develop a reliable correlation between the state of the fly and my swatting, by swatting when I observe the fly, I cannot exploit this correlation to influence the past. Deciding to swat on observing the fly already assumes that deliberation comes prior to decision. Given this, I will undermine any correlations to the past I attempt to exploit.

7. Conclusion

To conclude, we can use an asymmetry of records together with epistemic conditions on deliberation to explain why deliberation comes prior to decision. Given this, we can then appeal to evidential undermining (§5) to explain why we can't typically control the past. Agents with very unusual epistemic capacities may be able to deliberate after they act, and so effectively control the past. And there may be unusual circumstances (such as traditional Newcomb cases) in which we do control the past—cases in which correlations towards the past are not undermined by our deliberating on the basis of them. But in ordinary circumstances agents like us, able to form stable memories of past macroscopic events like

actions, won't be able to deliberate towards the past. We can't pick up on correlations between past actions and present decisions in a way that allows decisions to be epistemically undetermined. And given we deliberate before we decide, we cannot in ordinary cases exploit past directed correlations without undermining them.

By appealing to epistemic conditions on deliberation, we then have a good explanation for why we cannot control the past. Our epistemic capacities prevent past-directed deliberation, and given future-directed deliberation, we can no longer exploit correlations between our decision and events further in the past. This explanation could be used to defend Albert's original account of the asymmetry of control. Epistemic conditions and avoiding evidential undermining would then be further requirements on effective control of the past, of the type Albert invokes in responding to the Elga-Kutach and Frisch cases. *Contra* Price and Weslake (2009), appealing to features of agency is compatible with explaining the asymmetry of control in terms of entropy—particularly if explaining these features requires appealing to entropy. But the explanation I have offered here also stands alone. By appealing to plausible conditions on deliberation and an epistemic asymmetry that is more limited than Albert's own, we can explain why deliberation comes prior to action and why we cannot control the past.

References

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

———. Forthcoming. "The Sharpness of the Distinction between Past and Future." In *Asymmetries of Chance and Time*, ed. Alistair Wilson. Oxford: Oxford University Press.

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

Callender, Craig. 2004. "Measures, Explanation and the Past: Should 'Special' Initial Conditions Be Explained?" *British Journal for the Philosophy of Science* 55:195–217.

Dummett, Michael. 1964. "Bringing about the Past." *The Philosophical Review* 73 (3): 338–359.

Earman, John. 2006. "The "Past Hypothesis": Not Even False." *Studies in History and Philosophy of Modern Physics* 37:399–430.

Elga, Adam. 2001. "Statistical Mechanics and the Asymmetry of Counterfactual Dependence." *Philosophy of Science* 68 (3): S313-S324.

Frisch, Mathias. 2007. "Causation, Counterfactuals, and the Past-Hypothesis." In *Causation, Physics, and the Constitution of Reality*, ed. Huw Price and Richard Corry, 293–326. Oxford: Oxford University Press.

———. 2010. "Does a Low-Entropy Constraint Prevent Us from Influencing the Past?" In *Time, Chance and Reduction*, ed. Gerhard Ernst and Andreas Hüttemann, 13–33. Cambridge: Cambridge University Press.

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

Greene, Brian. 2004. *The Fabric of the Cosmos*. New York: Alfred A. Knopf.

Hampshire, Stuart, and H. L. A. Hart. 1958. "Decision, Intention and Certainty." *Mind* 67 (265): 1-12.

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

Horwich, Paul. 1987. *Asymmetries in Time*. Cambridge, Mass.: Massachusetts Institute of Technology Press.

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

———. 2012. "Decision and the Open Future." In *The Future of the Philosophy of Time*, ed. Adrian Bardon, 149–68. London: Routledge.

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

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

Lewis, David. 1979. "Counterfactual Dependence and Time's Arrow." *Nous* 13: 455−76.

Loewer, Barry. 2007. "Counterfactuals and the Second Law." In *Causation, Physics, and the Constitution of Reality*, ed. Huw Price and Richard Corry, 293−326. Oxford: Oxford University Press.

———. 2012. "Two accounts of laws and time." *Philosophical Studies* 160 (1): 115-137.

Penrose, Roger. 2005. *The Road to Reality*. New York: Alfred A. Knopf.

Price, Huw. 1986. "Against Causal Decision Theory." *Synthese* 67:195−212.

———.1991. "Agency and Probabilistic Causality." *British Journal for the Philosophy of Science* 42 (2): 157–176.

———. 2002. "Boltzmann's Time Bomb." *British Journal for the Philosophy of Science* 53:83−119.

Price, Huw and Weslake, Brad. 2009. "The Time-Asymmetry of Causation." In *The Oxford Handbook of Causation*, ed. Helen Beebee, Christopher Hitchcock and Peter Menzies, 414–43. Oxford: Oxford University Press.

Reichenbach, Hans. 1956. *The Direction of Time*. Mineola: Dover Publications.

Schick, Frederic. 1979. "Self-knowledge, Uncertainty, and Choice." *British Journal for the Philosophy of Science* 30:235−52.

Sklar, Lawrence. 1993. *Physics and Chance*. Cambridge: Cambridge University Press.

Taylor, Richard. 1964. "Deliberation and Foreknowledge." *American Philosophical Quarterly* 1:73−80.

Winsberg. Eric. 2004. "Can Conditioning on the Past Hypothesis Militate Against the Reversibility Objections?" *Philosophy of Science* 71 (4): 489–504.