The Free Actions of Glorified Saints

Richard Anthony Stuart Tamburro

PhD

University of York

Philosophy

June 2014

Abstract

This project examines whether we can consistently make two claims: i) God cannot prevent sin without destroying free will; and ii) in heaven, God prevents sin without destroying free will. It explores the conditions under which agents may be said to be acting freely, guided by consideration of the free will defence to the problem of evil. This involves considering prominent themes in the philosophy of action, and in the metaphysics of free will. It develops, and defends, a non-causal incompatibilist account of free agency. This is then applied to the theological problem of how God ensures the sinlessness of the saints in heaven, and why this could not be achieved prior to heaven.

Contents

Author's Declaration

I declare that this thesis is a presentation of original work and I am the sole author. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as references.

1. Evil, Freedom and Heavenly Agents

Introduction

This is a project in philosophical theology. As such its purpose is two-fold, first, to illuminate a theological puzzle by applying the tools of analytic philosophy, and second, to explore the relationship between commitments in a philosophical puzzle more clearly by setting it in a well-defined theological framework.[1] There will be limitations on what we can achieve: we will not provide anything like an exhaustive argument for what all of our philosophical and theological commitments should be, although our exploration of the issues will, of course, examine some of the reasons we might be led to our commitments. The approach we will adopt is more counterfactual – if we are committed to this one thing, how does that commit us in another area, so it involves us in some substantial assumptions. The details of positions we consider will still matter. We will be interested in establishing the possibility of coherence between claims, and the details of those positions are the basis of demonstrating that there is a coherent position to be had. However, there may be some objections to the positions we will consider that we will not be interested in entertaining given our assumptions. The purpose of this first chapter is to map out the theological puzzle that motivates this project, the philosophical issues it raises, and to make explicit the key assumptions and limitations. In its briefest form, the puzzle we will address is one of consistency between two claims, which appear prima facie to be contradictory.

i) God cannot prevent sin without destroying free will;
ii) In heaven, God prevents sin without destroying free will.

The first statement arises from a response to the problem of evil, and the second arises out of theological considerations pertaining to the state of human beings in heaven. Our central question is whether there is a coherent theory of what it is for humans to be free agents that enables us to make both of these claims without contradiction. Developing an answer to this question will enable us to clarify what God can and cannot do concerning his interaction with free beings through better understanding the nature of what it is to be a free agent.

[1] An approach well elucidated by Crisp (2009) and (2011). Also note that we will limit ourselves to the Christian theological tradition.

The problem of evil and the free will defence

The free will defence is a response to the problem of evil. The problem of evil is a broad problem with many aspects, but the aspect with which the free will defence engages is why the evils that are in the world obtain, rather than not. We will not argue that the free will defence is a successful defence, though it is widely taken to be so, but will be asking how we should proceed if we are committed to making this sort of defence. As we will shortly see, the free will defence does not just commit us to free will in any sense, but in a particularly libertarian sense.[2] The defence proceeds along the following lines: given that there is some evil in the world, and that evil is a bad-making thing, when we ask why it is there, we would like to be able to answer that it is not because of God. Whatever the reason for evil is, it had better not be God. To have God be responsible for evil would be problematic, since we do not expect a wholly good God to produce anything other than good. So, in order to escape this conclusion we say that the reason that evil is in the world is not, ultimately, because of God, but is because of the actions, or choices, of free creatures. God could not prevent his free creatures from sinning (or falling) without destroying their capacity for freedom. So, when investigating where to attribute responsibility, the buck stops not with God, but with free creatures.

Asking where the buck stops is a useful turn of phrase for seeing why one might be concerned about evil in this way. It depends on a particular connection between causation and responsibility. We do not have space to consider in detail what responsibility consists in, it is a particularly complex issue, but there is one way of looking at responsibility, by no means the only, that we will assume in order to get the free will defence going.[3] Part of what it is to identify when someone is responsible is, plausibly, to identify who is the cause of the event or

[2] Although the terms have widespread consensus on usage, note that in our discussions, incompatibilist denotes someone who thinks that free will is not compatible with causal determinism, while a compatibilist thinks that it is. A libertarian is an incompatibilist who thinks that there is actually free will, and is thus committed to the falsity of causal determinism. Causal determinism, or often just determinism, is not as simple to define as it might appear, but is captured by the idea that given some complete description of the world at a time, and given the laws of nature, there is only one possible future state of the world for any given future time. Determinists are committed to the view that there is no causal slack, or leeway, in the causal evolution of the world. Not many people are committed to this strong form of determinism, but we will discuss variations of the issue in later chapters. See Hoefer (2010), Steward (2012) and Vihvelin (2013) for a more thorough discussion, especially of varieties of determinism, and the importance of how we conceive of the 'laws'.

[3] This assumption is not arbitrary, since this claim about responsibility seems to be an intuition that drives most proponents of the free will defence, but we will not argue that this is so, though it is interesting to note.

state that is morally praiseworthy or blameworthy.[4] If Tom stabs Dick, killing him, then Tom is responsible for something bad happening, and is an appropriate recipient of blame, punishment, shame, and so on. However, if Tom should protest that he never intended to stab Dick, and was only trying to cut his steak, when Harry grabbed his arm and moved it to stab Dick, then we are able to quickly conclude that Harry is responsible for what has happened to Dick, and not Tom. There are a complex array of relevant issues we might want to know about, not least Tom, Dick and Harry's intentions – after all perhaps Tom told Harry to grab his arm, or there was an element of self-defence. But avoiding variations on the basic scenario that may introduce these considerations, Harry's being the cause of the event of Dick being stabbed makes Harry responsible. We do need to be a little careful with the word 'cause' here. Harry might say he was caused to grab Tom's arm because of something Tom (or Dick) said, or that he would never have thought of doing anything had Tom not been holding a knife, and that one of these features of the conditions Harry found himself in was the cause. Harry might even go so far as to say that Tom knew about Harry's temper, and had ample opportunity to intervene, and that Tom's failure to do anything resulted (which implies a causal contribution) in Dick being stabbed. Perhaps Harry is a psychiatric patient prone to grabbing people in fits of rage, and perhaps Tom is the psychiatric nurse responsible for overseeing Harry – in this case it may be Tom's negligence that could be said to have caused the stabbing of Dick.[5]

It is useful to distinguish between agential and moral responsibility. Agential responsibility picks out the agent whose action brings about the state of affairs that we are considering. This distinguishes agents who are involved in events as instrumental causes, from those who are involved as efficient causes. To be agentially responsible an agent must be the efficient cause of an action.[6] Moral responsibility identifies the agent whose exercise of their agency, through action or inaction, proximal or distal to the bad-instantiating event, makes them responsible in the sense that they are justifiable objects of blame and shame.[7] So where Harry is a psychiatric patient, he is agentially responsible for the stabbing of Dick, but perhaps, depending on his condition, not morally responsible. In the same scenario Tom is not agentially responsible for the stabbing of Dick, but may be morally responsible, especially if he was relevantly in the know about Harry's

[4] This is a contentious issue, but forms part of our assumptions. For a brief consideration of how a libertarian might go about defending this claim see the dialogue between Kane and Fischer (2007).

[5] For a brief discussion of omissions as actions see Bach (2010), and Alvarez (2001).

[6] This is more than a little vague, but we will discuss this issue in detail later. It is enough now to label the difference for the purpose of explaining the free will defence, if not to analyse it fully, and to await later discussion.

[7] So it is, contra Strawson, satisfying certain conditions that makes one responsible, see Strawson (2008).

behaviour. Of course Tom may have stabbed Dick in a well-considered plan of hatred, and he could be agentially and morally responsible. Or the situation may be very similar to the one involving Harry as the psychiatric patient, except that Harry is not a patient, and is only a well-known friend of Tom. We can imagine Tom and Harry getting into an altercation with Dick, and Tom threatening Dick with his knife, but Harry grabbing the knife and stabbing Dick with it. Now Harry is agentially and morally responsible, but Tom may also be morally responsible.[8] Notice that we often have differing intuitions about *how much* blame to apportion to agents in these different circumstances.[9] Tom's premeditated, deliberate stabbing of Dick in hatred seems to be a moral responsibility that involves his being more blameworthy than in the case where he is negligent, and allows Harry to grab his arm and stab Dick. I will not argue that this judgement is well founded here (although I think that it is), but do think that it is a judgement that advocates of the free will defence should espouse, since it does seem to motivate the strategy of the argument.

Consider now God's relation to responsibility. The situation involving God and the bringing about of evil can be taken as partially analogous to Tom, Dick and Harry cases, depending on how it is conceived. God takes the role of Tom, we, the creatures, take the role of Harry, and the stabbing of Dick, is the evil. In something like a process theology conception of creation our assessment will be like the case of Harry stabbing Dick, without Tom knowing it would happen, and thus not being in a position to foresee and prevent it. For the process theologian God's knowledge is like our natural knowledge, and comes via his experience of the world, so his providence is limited in some ways.[10] God does not foreknow that evil will happen, though he may know it is a possibility, and in a sense foresee it as likely (in the same way I might be able to foresee that my children's fighting will all end in tears). Analogously Tom may know that Harry is capable of stabbing people, but without some direct reason to suspect Harry of any malevolent intentions, he has no reason to hide all the knives from Harry. So for the process theologian to deal with the aspect of the problem of evil we are interested in, all they need to do is make God relatively surprised at what creatures end up doing so that God is not placed in a position of culpable negligence concerning his providential actions in the world, since God is not about to possess sovereign providential control. In the limiting case God knows nothing of his

[8] He is agentially responsible for the threat, but not the eventual stabbing.
[9] Though I hope the examples are ones in which our intuitions about who is blameworthy are in general agreement. There is considerable disagreement sometimes about what the agents are to be blamed for, but this is a complication we will ignore for now, and leave to another project.
[10] Actually process theologians come in many shapes and sizes, and would not all say this, but it characterizes one view worth our consideration.

creatures first actions until they are performed and he begins to get to know them, and thus bears what seems to be no moral responsibility, since he can exercise no providential control over this, in the same way that Tom would not be blamed if Harry was a stranger to him, and gave Tom no reason to suspect immanent foul play. Process theologians can achieve this by making creatures autonomous agents, with the ability to act in a way that is not controlled by God, and is disentangled from his influence. God's providence over his creatures is then reactive.

So the problem of evil is, at least in this regard, not much of a problem at all for process theology. However, the free will defence was developed, in the main, by those with a desire to maintain God's providence in a more conventional sense, so that God actually possesses infallible foreknowledge, and complete, sovereign guidance and control over the unfolding of the world, and cannot act (or be) surprised when his creatures begin to sin. So we need to consider a free will defence scenario,[11] one more like the case of Harry the psychiatric patient. God may know that we have a weakness that makes us prone to do things that are bad, or if not prone, admits of the possibility of our so doing. But God does not intervene to prevent our acting in a bad way, in an analogous way to Tom's failure to intervene to prevent Harry stabbing Dick. But now comes the crucial part of the defence. If Harry is morally responsible for what he does, and he is agentially responsible as well, then he bears a greater responsibility than Tom, who is not agentially responsible for Harry's act. We made Harry a psychiatric patient, who may plausible fail to be morally responsible if he lacks certain cognitive capacities, so we need to alter the situation a little. Harry may still be under Tom's care, but has the faculties necessary to be responsible (in both senses) for what he does (it does not matter to this part of our argument what these are) – perhaps he is under surveillance, or is undergoing a psychiatric evaluation. So with God's creatures, they possess the faculties necessary for them to be responsible, and are thus agentially and morally responsible for their bad acts. God however, remains only morally responsible, for permitting through failing to intervene, or to providentially prevent these acts, and the blame attached to this seems to be less severe.

We need to consider this free will defence scenario in contrast with a compatibilist scenario to complete the free will defence. In the compatibilist scenario the person that performs the bad act was determined to do so. So the situation is more like Harry grabbing Tom's arm, and causing him to stab Dick. Tom was only instrumentally involved, and did not seem to be blameworthy, or even if the recipient of some blame, was much less blameworthy than Harry, who

[11] As apposed to the process theology scenario.

bears the brunt of moral responsibility. So in our analogy the roles are different: we, the creatures, are the ones who are 'grabbed' like Tom is, or determined to do as we do, by the chain of past events which lead up to this moment. This past chain terminates in something that bears the agential responsibility (or is the efficient cause) for what follows, and in our case, this is God. So in the compatibilist scenario creatures are only morally responsible (if even that), whereas God is both morally and agentially responsible.

Now we can see how the free will defence scenario mitigates the problem of evil compared to the compatibilist scenario. Remember that where someone is both agentially and morally responsible, they bear a greater proportion of the blame than the person who is morally responsible, but not agentially responsible. Now, in the former scenario God is only morally responsible for what he permits to occur, whereas in the latter God is agentially responsible for the occurrence of everything, and is also morally responsible for everything that occurs. The free will defence does not deny that there are some further questions outstanding regarding God's responsibility for evil, however it does divide up the blame in a way that shifts the focus to creatures, and away from God. Notice that this argument does not depend on the claim that agential responsibility is required for moral responsibility, or the claim that in the compatibilist scenario creatures are not responsible at all (which amounts to the same thing). Although many people have attempted to argue this, it has proven a decidedly difficult claim to convince compatibilists of. Compatibilists have claimed to be able to analyse our concepts of moral responsibility in terms of instrumental responsibility (where agents are 'moved' by determining factors, like Tom is by Harry), with the addition of further criteria that 'involve' the agent in what occurs in some special way.[12] Although I do think that there is something to the claim that there could be no moral responsibility without the possibility of exercising agential responsibility, it is not something we need to argue, and this is a strength of this formulation of the free will defence.

Agents, aseity, and freedom

So the free will defence hinges on creatures being able to be agentially responsible for some bad events, occurrences or actions. We noted that we might begin to understand this intuition by the distinction between an efficient and instrumental cause. We also noted that we can approach the issue by asking where the buck stops as we attribute responsibility. Whatever it is for a creature to be agentially responsible, it must be a condition that prohibits us from reanalysing

[12] For an introduction to the vast literature on this issue see Fischer (1999) and Eshleman (2014).

the situation in a way that shows the agent to be instrumentally involved, and reveals God (or the past) as the one that is agentially responsible. This means that creatures will need to be able to act in a way that is considerably independent from determining factors. If we consider some real instance of evil, or sin, such as my deliberately stabbing someone, then *I* am at fault for stabbing someone. It will not do for me to say that I was forced into it, or was so constituted by my genetics, or God's providence, to make my stabbing inevitable. The buck stops with me.[13] Although there may be many facts about the past, about influences over my present constitution, and about God, all of which may have made it possible for me to sin, the making the sin actual was something that I did. I may not be agentially responsible for the potential for sinning (ultimately I cannot be responsible for my own birth), but moving from potentiality to actuality was something that was within my power.

Since we are using an Aristotelian way of talking about this, we might be reminded of his 'unmoved mover' – that which moves without itself being moved.[14] When Harry grabbed Tom's arm, Tom was moved. But if Tom moves by himself, without any prior external conditions necessitating his so doing, then he moves himself. This captures something of what Christians want to claim about God, that he initiates activity in the universe, and that all causally related chains of events terminate in God, but that God's own activity is not dependent on anything prior to, or external to, God. In God's case he is entirely 'unmoved', there being nothing other than himself to even structure the possibilities of his movement. In Tom's case he does not enjoy the same sort of complete independence. God is *casua sui*, having the ground of his own existence, and activity, in himself. Theologians talk about God's aseity (from the Latin *a se* – from himself), his independence, or rather *self*-dependence, when we explain his being and activity.[15] Proponents of the free will defence want to say something similar about creatures. The idea captures one similar to the claim we are trying to make about agential responsibility. Although creatures do not need to possess aseity in the same way that God does,[16] they do need to possess the possibility of aseity regarding at least some of their actions. It is these moments of creaturely aseity that form the basis of being able to make the buck stop with creatures rather

[13] Ultimately we will be interested in the instance of evil involved in the fall, both that of Adam, and of the Devil. When Adam took the fruit and ate it he made that instance of evil obtain, but God did not make Adam take the fruit.
[14] See Chisholm (1964).
[15] See Strong (1907) and Grudem (1994) for a modern theological discussion of aseity, where the focus is normally on being rather than activity. The scholastics have a lot to say on the subject, especially Aquinas, who develops the Aristotelian idea we have noted into arguments for God's existence.
[16] Indeed that claim would be quite counter to Christian orthodoxy, in which all is dependent on God in some way.

than God when we are considering responsibility and the apportioning of blame. Sometimes creatures are, in a sense, unmoved movers, and although the possibility of their exercising this power may depend on God, and even the possibilities over which this power can range may depend on God, the exercise of the power is one that the creatures are agentially responsible for.[17] It is plausible to claim that the possession of this aseity is part of what constitutes being made in the image of God.

Now this is not much of an argument that creatures do, or should, possess this characteristic. Indeed, we have merely tried to describe it, and identify it. However, it is required for the free will defence to work. So part of what it is to be committed to the free will defence as a response to the problem of evil, is to be committed to at least some form of creaturely aseity. This aseity is, hopefully, recognisable as something we sometimes term freedom. I have deliberately shied away from couching our previous discussion in terms of freedom, or free will (even though the concept is implied by the name of the defence) because it is a term that is used to refer to a wide range of concepts. Indeed, forms of the free will defence have been suggested that are supposed to be compatibilist friendly, so care is required.[18] However, freedom is an appropriate term once the sense of freedom we are interested in has been clarified. The free will defence requires an account of freedom that is libertarian, since a compatibilist notion of freedom will not allow us to make the sort of distinction we need to make to move the argument forward. Agents must be free to act in a way that leaves them agentially and morally responsible. We will not argue in detail that agents do possess aseity – it comes with our assumption of the free will defence. However we will be concerned with what this aseity looks like once unpacked. Since our target is what it is for agents to perform actions that are free in this sense, we will not approach the question of what it is for agents to be free in terms of what it is for agents to be morally responsible (as is often the case in the free will literature), but instead will approach the issue from the direction of what actions are, how they relate to agents, and what it is for an action to be free in the relevant sense.[19]

---

[17] Obviously if the possibilities over which this power ranges are all evil, then creatures find themselves in somewhat of a fix, one that they may justifiably complain about. So it is important that as well as the possibility for sin, creatures have the possibility of righteousness – to good as well as evil. We will say more about the range of possibilities in due course.

[18] See Cowan (2011), and a discussion of why compatibilist versions fail by Pawl & Timpe (2013).

[19] In fact, as we will see, for an agent to be agentially responsible actually involves several senses of freedom, all of which can prevent an agent being in the right sort of relation to a bad action to be the place where the buck stops.

What is so good about freedom?

We should note though that an assumption of the free will defence involves us in more than just this. We discussed aseity in the context of a version of the free will defence that maintains a strong sense of divine providence. Process theologians may be able to claim that God permits creatures to sin, but that this is an unfortunate and unforeseen accident,[20] but the free will defence is at its most interesting when providence is squarely on the table. This is especially so given the problem we are interested in, for if God is surprised by his creatures' actions, the idea that in heaven God secures a state that necessarily excludes sin, may be a problem. However, the two claims of aseity, and of providence raise another puzzle of coherence. How is it that the universe can be under God's sovereign guidance and control yet we, part of that universe, remain free? We do not have space to unravel this puzzle, though our later analysis of freedom may improve the prospects for usefully attempting to unravel it at a later date. We will be assuming that there is a libertarian friendly solution to this problem. Although it is not without its challenges, I suspect that Molinism will be of assistance in this regard.[21] This is a well-developed attempt to reconcile freedom and providence, and so this assumption is at least not without some warrant.

But a more pressing issue for us to consider is about what else must be true of freedom, if the free will defence is to succeed. Although our being agentially responsible shifts the focus of moral responsibility for the actualising of evil onto free creatures, there still remains a question about the responsibility for creating the possibility of creatures doing so in the first place. Given that God possesses providential control over what will happen if this possibility is created, it seems that God is morally responsible for permitting evil. It is a part of the free will defence to claim that God cannot prevent free creatures from exercising their freedom in evil ways without also destroying their freedom itself. After all, our definition of aseity explicitly prohibits determination of its exercise by anyone other than the possessor of that aseity. But it is possible that God could have not made creatures free, and so circumvented the possibility of their sinning. There is another sense in which God might be thought to be permitting evil, and that is through his failure to intervene when evil is about to occur. However, one can respond that God has, in fact, intervened on multiple occasions to prevent or

[20] Though again, many would dissent, and not want to claim exactly this.

[21] The central problem seems to be the grounding objection, but it seems to me that the grounding objection will be a problem for any account of divine foreknowledge given anything like presentism. If presentism is not the correct metaphysics of temporal being, then the problem looks eminently soluble without recourse to counterfactuals of creaturely freedom, but then, I suspect, libertarianism may need to be given up or radically revised as well. I do not think that these issues will affect our discussion, but the possible connection between these ideas is noteworthy. See Perszyk (2011) for a recent volume on the state of the Molinism debate.

remedy evil, most notably in the incarnation. Even then though, those who see evil as a problem are likely to wonder why God does not intervene more often.[22] This is a different aspect of the problem,[23] and not one that we will consider, though it is not an aspect that theists can ignore. The point we need to notice for now is that in the sense in which God is responsible for permitting evil through creating the possibility of evil by creating creaturely freedom, God requires some justification for so doing. Although 'risk' is a somewhat inappropriate term given God's providential guidance and control, asking why freedom is worth the risk of all the evil that has now obtained, gets to the heart of the matter. Since God knows all the evil that will befall creation if he makes creatures with aseity, permitting this to happen, that is, making it possible for creatures to become responsible for so much evil, can only be morally justified if doing so secures some good which outweighs all of the evil.

Therefore, freedom must be a great good, and considering the prevalence of evil, and its severity, it must be a very great good indeed.[24] There are a number of responses to this issue. Some, like Plantinga (1974) focus on the fact that we are not in a good epistemic position to know what reasons God may have for permitting evil, and that as long there is not reason to think that a justifying reason is in principle impossible, we are left able to conclude that the problem of evil is not an insoluble problem.[25] However consideration of what the form of these justifying reasons could be reveals an issue for the heavenly end of the problem we are considering. Justifying reasons for freedom will focus on freedom having some intrinsic value, or on having some instrumental value. If it is the intrinsic value of freedom that justifies the 'risk' of all of the evil, then the presence of freedom must be a very good-making property indeed. Now given that the time creatures spend in heaven is eternity, whereas the time during which creatures have been, and will be, able to exercise freedom on earth before then is relatively

[22] It can be argued that God is under no obligation to intervene, given that creaturely freedom includes bearing the consequences of our actions. Consider Harry as the psychiatric patient again. If Tom is not charged with Harry's supervision, but is instead a stranger, then it might well be that Tom could have intervened in some way, but was not under any duty to do so – if he does it will be a supererogatory act. However, given God's special relationship with creatures he creates in Christianity, and the duty of care he espouses himself to be under, this does not seem to be a tactic that will work for a Christian response to the problem.

[23] Sometimes called the probabilistic problem of evil, but this term denotes a wide range of issues as well.

[24] Some have questioned whether libertarian freedom really secures any good not secured by compatibilism. See Mele's Diana fable (2006:8 ff.), there are many other compatibilist accounts of these 'goods', but we do not have space to critique them all.

[25] In his terms, the problem of evil is not a defeater for theism. This marks Plantinga's approach, and other like him, as offering a defence that seeks to explain why the problem is possibly soluble. A theodicy tries to explain what the solution actually is, and what the reasons actually are.

short (by comparison), we can expect that if freedom is a great-making property that is valuable, it will be present in heaven as well as on earth. There are two possible ways of reaching this conclusion. First, the freedom on earth may outweigh the evil, and thus balance the books, but then, if freedom is valuable enough to achieve this then heaven would be a poor place without it. This amounts to the claim that we expect heaven to contain very great goods, and that if freedom is such a great good, its absence in heaven would be a bad thing (and heaven cannot be bad in any way). Secondly, perhaps the good of earthly freedom does not outweigh the evil, and so its continued presence in heaven is needed to add to the scales, and outweigh the evil.

What if freedom is instrumentally good? Well the situation may be very much the same, and if freedom is instrumentally required for great goods, if we want those great goods in heaven, then we will need freedom in heaven as well. This may be to outweigh the evil, or may be because heaven is such that it should contain great goods. However, it may be possible that freedom is required to achieve some great good, or to bring about something of value, but that freedom is not required to maintain that good. In this case it would be possible to balance the books without requiring freedom in heaven. Consider a plausible example: perhaps I need to exercise freedom in order to achieve the valuable state of my possessing salvation, but once achieved I do not need to keep on exercising freedom in order to maintain this state. If this is the case then I may need the opportunity to exercise freedom on earth, but not in heaven. But the value of this good continues into heaven and so we stand a good chance of justifying the permission of evil, and do not miss out on this good in heaven. Now if this were the case, and supposing this good to be the only good we need freedom for, then there would be no problem of consistency for us to solve, because our initial claim, ii) in heaven, God prevents sin without destroying free will, would be false. It would be open to God to prevent sin and destroy free will at the same time – once the good of salvation had been secured. Now this would make our proposed project an uninteresting one, and in order to have an interesting theological puzzle for us to get our philosophical teeth into, it is tempting to simply assume ii). However I think we can say something about why ii) is true; about why the proposed example fails, and about what goods we might need freedom for, and why they are not of a kind that can be maintained without the continued possession of freedom (note that we only need one freedom involving good to be like this to counter the above move). It will be difficult to argue in as much depth as we would like given space considerations, so our discussion here will be a little speculative, but a short discussion will illuminate issues we will need to refer to later.

To establish the truth of ii) we need to see how it is possible for God to prevent sin, and this will be the concern of later chapters. But we also need to see why it is desirable that God is bound by ii), in other words, why it would be a bad thing to destroy free will in heaven. I will argue that there must be goods that can only obtain if there is a concurrent capacity to exercise freedom. To begin thinking about this consider the example of salvation again. Is salvation an intrinsic good, or not? It appears to be instrumental. It is good that we are saved, if we are thereby rescued from suffering. It is good that we are saved if we thereby get to enjoy peace, joy, love and so on. It is the presence of 'thereby' in these sentences that indicates the real reason that salvation is valuable – it is because of what it gets us. Let us just label some good heavenly possibility that is available to us when saved as *G*. Now it may well be that having exercised freedom to make *G* something that is available to us, we do not need to exercise our freedom further to maintain the availability of *G*. However, the key question is whether our interaction with *G*, or our *G-ing*, is something that requires freedom or not. If *G-ing* does not require freedom, and if *G-ing* is the great good, the valuable thing that justifies the risk, then we might ask why God did not just create unfree creatures who *G*. If our ability to *G* does not depend on the capacity for freedom, then it looks quite difficult to see why it would not be possible for God to make creatures that are not free, and hence cannot sin, but can still instantiate the good of *G-ing*. If *G-ing* does require freedom, then although questions about salvation may raise interesting issues about how we can exercise freedom valuably, there is no reason to suppose that it would be good for God to destroy freedom in heaven, in fact his so doing would destroy the possibility of a great good. So if we are committed to the free will defence, we have, prima facie, a strong reason to think that the good that justifies the creation of free creatures is one that will also be present in heaven, and will require the continued presence of creaturely freedom.

So what great good could be a plausible candidate for a good that justifies evil and requires freedom, but also requires freedom if it is not to be lost? There are a number of possible goods whose dependence on libertarian freedom has been suggested. For example Kane (1998:79-101) suggests creativity, self-creation, desert, moral responsibility, interpersonal reactive attitudes, individuality, dignity, life-hopes, love and friendship and objective worth.[26] We will not examine all of these – that would be a substantial undertaking. Compatibilist friendly accounts of most of these characteristics have been offered, but libertarians persist in claiming that there is an intuition about the good in these characteristics (or some of them) depending on the possession of libertarian freedom. This is not

[26] That libertarian freedom is required for these goods is not uncontroversial, but that would require another chapter to argue, or more. See Clarke (2003) and Pereboom (2001), especially on love.

particularly convincing in and of itself, but it does indicate that there is a wide array of options that the free will defence could explore. Let us focus on one issue, that of relationship. A relationship with God, one characterised by love, knowledge of God, trust and worship, is prominent in Christian thought as a valuable good, and even an incommensurable one. Now, being in a relationship is a state, but it is a state whose obtaining depends on the exercise of various interpersonal interactions, such as loving someone, trusting someone, communicating with someone, and so on. We should note the distinction between a good residing in a state, or in the exercise of some power, or activity. If there is any prospect of discovering a candidate for *G* whose maintenance does not depend on freedom, it seems likely to be a state. It is plausible that entering into a state could require an exercise of freedom, but that once entered into, no further exercise is required. We are familiar with this through mundane examples: my entering the state of being sat at my desk requires some action to get me there, but once sat there, I do not need to keep myself there by some active exercise of my powers – I could even fall asleep and remain there. Similarly my entering the state of being married requires my active involvement in saying my vows, but once the wedding is over I do not need to keep on saying them.[27]

But many of the aspects of relationship are not states, and involve the active exercise of my ability to act freely. Consider one aspect of love in a relationship with God, namely to be a worshipper of God (which involves the expression of love towards another[28]). There are many states relevant to this, such as my possessing right beliefs about God. However what it is to worship God, is not only to know that God is great, but also to respond to that knowledge in an appropriate way. Now responding is reactive, and if there is to be any value to my reaction, then I must be reacting because of the value I discover in God. I cannot be forced to play my heavenly harp for example. But if I do not have freedom, then it is not possible for me to truly react to this value. If I am not free, and the things that happen involving me, which are supposed to constitute worship, are caused to happen by God, then they are not appropriate candidates for being acts of worship. This same intuition is sometimes expressed in the idea that you could not programme a robot to truly love you, since having caused all 'loving reactions' they would not really constitute acts of love at all. Now, this is a difficult intuition to justify, as it deals with the knotty problem of acting *for* a reason, which is an issue on which libertarians and compatibilists are prone to disagree, and we will give this some discussion in a later chapter. But I hope that brief consideration of

[27] This would be marriage as a legal institution, a marriage considered as a form of relationship may well require me to continue to live my vows.

[28] This is a pertinent example, since although one might claim that love exists in heaven, if it is not expressed and enjoyed, we would seem to have missed the point of love somewhat (the thought is certainly suggestive).

this claim does identify that there is something about being a worshipper that requires libertarian freedom.[29] Lowe has given one of the clearest articulations of this thought.

> For a substance to act in a certain way because it was *caused* to act in that way … is not for that substance to act rationally, or genuinely 'in the light' of any reasons it may have for so acting. By so acting, a substance may act in a way that is *in accordance with* certain of its reasons for acting, if it has any, but it cannot be acting *for* any of those reasons, given that it was caused to act in that way. To act for a reason is to act in a way that is responsive to the cogency of certain considerations in favour of one's so acting – and this is incompatible with one's being caused to act in that way, because causal processes bring about their effects with complete indifference to the question of whether this effects have cogent considerations in their favour. To act in the light of one's reasons for acting in this or that way, one must, then, be able to *choose* so to act… [italics original] (Lowe, 2008:156).

If the active exercise of my ability to interact with something good is to have value, I must be responding, here and now, to the goodness of that thing as I directly appreciate it. It is not enough to say that I encountered some direct appreciation of something's goodness in the past, and that this can be the basis of what I do now, as if what I am now doing is part of an extended response, or reaction, initiated by some freedom involving event in the past. Imagine if a husband claimed this about his wife, with the marriage now considered relationally, not legally. To say his current avowal, or expression, of love towards his wife is genuine because he did, after all, fall in love with her in the past, would not be counted genuine. What she wants to know is whether he loves her now. So although we talk of relationships as a state we are in, the continuation of a relationship depends on the continued free and active exercise of agents in the relationship towards one another in various ways. In the case of a relationship with God we considered worshipping, but the principle extends to many other aspects of participating in a relationship. We will not explore what these different aspects are now, since a philosophical account of relationship would be a significant detour. We only needed one plausible example of a *G* that depends on freedom for it's persistence, and we have that in worshipping. However I hope that is not difficult for the reader to imagine how this thought could be extended to other aspects of relationship with God. So we have a reason to think that at least some of the goods associated with justifying the permission of evil depend on freedom not only for their inception, but also for their continued persistence, and

[29] That the possession of an objective practical reason depends on the freedom to do otherwise is also discussed by Haji (2012), though his focus is in the context of moral responsibility. Wallace (2004) discusses the connection to normativity and the will, which may suggest that those with a Kantian intuition will agree with this claim.

given that we expect these goods to be present in heaven, we have good reason to think that the same sort of freedom required to make the acquisition of these goods possible is also present in heaven. This gives us a defence of claim ii).

Active engagement in great goods

However it is not only the presence of freedom in heaven that matters. Susan Wolf has a phrase she has used to describe what it takes for agents to have meaning in their lives, and although she is not a libertarian, and was not directing her comments at the issue of freedom, it is apposite to our discussion. To have meaning, or significance, in their lives, agents need to be appropriately connected to things that are good or valuable. They need to have 'active engagement in projects of worth' where 'subjective attraction meets objective attractiveness' (Wolf, 1997:209, 211). We have considered the need for agents to possess active engagement; this is a matter of their having the capacity to freely engage in a certain interaction. But it is not enough to just stipulate that agents in heaven must possess freedom, because if the exercise of that freedom is to be of great value, or involve the instantiation of great goods, then *how* that freedom can be exercised is also critically important. For example if I am free in the relevant sense, but can only exercise this freedom over the decision of whether to begin walking with my left foot or right foot first, then it would be difficult to see how this exercise of freedom could instantiate, or bring about, a great good. Now, given our discussion, and the examples we have used, it should be clear that there is scope for freedom to be exercised in more significant ways than this. But since this is such an important point, it is good to be explicit. Whatever these projects are, that heavenly agents engage their freedom in the pursuit of, they must be projects of great worth. In order to be a project of great worth, the project must instantiate some great good. But in order to be actively involved in the pursuit of that project, an agent must exercise their freedom in choosing that project, or in carrying it out. Therefore agents must be involved in an evaluative judgement about the normative goodness of whatever possibilities there are for them to exercise their heavenly freedom regarding. Agents must make a judgement about which projects are 'objectively attractive', which will be a basis of their practical and evaluative rationality concerning their heavenly activity, and given the nature of heaven, we would expect agents to make a correct judgement about which projects are worth pursuing.

Now, herein lies a possible problem: if there is one correct normative judgement about which project is most worthwhile, and if failing to be subjectively attracted to that which is most objectively attractive is a sin, then agents only seem

to have one possible course of action open to them.[30] Now, does this mean that heavenly agents are so constrained regarding what they do, that they are no longer free? Well, as long as the subjective attraction is not caused, we may be able to satisfy the conditions for agents being freely engaged by the good and valuable.[31] However quite how God will ensure that agents always do the most valuable thing without causing what they do, or choose, in some way looks to be a difficult puzzle. But, of course, this is one of the questions we will be looking at once we have a clearer idea of what it takes for an agent to act freely, and how God can interact with agents without destroying this capacity. But even if it is possible, there is something intuitively valuable about the capacity for freedom being exercised in making choices, and this is often especially so for libertarians. We also noted that we expect the freedom-involving goods in heaven to require the continued exercise of freedom. Perhaps though there is scope for there to be choice among alternatives that are of equal value, for example, whether to sing bass or tenor in the heavenly choir.[32] But this begins to look like the choice of whether to walk with my left or right foot first, a choice that appears to be insignificant, even if where I am walking to, or why I am walking, is significant. If the exercise of freedom is so constrained as to preclude choices that are between options that are significant, then the presence of freedom in heaven seems to lose its appeal.

The intuition that there is a problem here is sometimes captured by children (and adults) who worry that heaven seems like a place that will be rather boring.[33] We seem to place a value on freedom as not only the freedom to be actively involved in *the* project of worth, but also in *which* projects of worth. It matters which goods we pursue, and how we pursue them. This is a theme echoed in Kane's list, and draws on features of agency such as self-creation, creativity, and

---

[30] This does depend on a plausible ethical principle that we are obligated to pursue the best goods we can. This principle is not without detractors (and it is particularly difficult to see how this fits with the idea of supererogatory acts). However, consider God, who in an Anselmian conception of a perfect being, must be he-who-does-that-than-which-no-better-an-be-conceived. This sort of conception of God is an instance of the application of just this principle, and since human agents are called to 'be holy as God is holy' this does provide some reason to think that we are also bound by this principle. This is far from a conclusive argument, but at least we have acknowledged that there is an important ethical question here. (Note that this observation also demonstrates that our discussion of whether heavenly agents possess freedom will also have some bearing on whether God is free. We will not discuss this question, but it would be an interesting application of the analysis we will undertake.)

[31] This touches on whether alternative possibilities are required for freedom, an issue we will assess in more detail later.

[32] This is a rather silly example, and a caricature of what heaven might be, but we will discuss heavenly choices in detail in the last chapter, and it suffices for an example.

[33] Walls (2010) connects this to Williams' Makropulos Case (in Fischer (1993)). Though we will not tackle this issue directly, our discussion can be applied to Williams' claim that eternity is a bad thing.

individuality. I am going to call a choice significant if it engages an agent's capacity for freedom between alternatives that engage the agent's evaluative faculties in an assessment that recognises more than one good option, and if the choice matters to the agent because in making that choice they are determining what goods they want to instantiate, associate themselves with, and to prioritise – the choice reflects their power to not only react to the goods around them, but to exercise their freedom over who they are, and the goods that constitute their character, interests, loves, and so on. Moreover, heaven must contain significant freedom, if heavenly freedom is to be a great good. Put another way, we want to know whether the need for alignment with objective attractiveness allows agents any slack concerning how they are subjectively attracted, and to what. This is a little stipulative, as we have not considered all the details of this definition. However it is not wholly stipulative, since we have discussed the need for a real exercise of freedom in heaven if we are committed to the free will defence. We have also sketched some of the concerns that might drive the intuition that freedom is valued for its significant exercise.[34] So if we are assuming that heaven must contain significant freedom, then we have, at least, weakly justified this claim. But there is another reason to include the claim that heaven must contain significant freedom as a desideratum, and that is that it makes this project more interesting. Although it is perhaps not necessary to do so, we have set the bar intentionally high in seeking whether heaven could contain the highest quality of freedom we can justifiably imagine or expect, and yet could still be a place without the possibility of sin. This is important because the theologian especially will be particularly interested in exploring how agents can express themselves, and relate to God and each other, in heaven. If we can construct a framework for understanding how the philosophical issues constrain these possibilities, then so much the better.

The sinlessness of the saints

We have used the idea that heaven must be without sin, but have not said much about this. Heaven must be a place without sin, because sin is something that separates creatures from God, and in heaven creatures are with God, and in the most intimate contact with God.[35] The salvation Christianity espouses is one in which our past sins are separated from us, enabling our entry into heaven and

[34] A discussion of the intuitions responsible for libertarianism would be a lengthy detour, but I suspect that the intuitions which are shared by libertarians, and by proponents of the free will defence, would lead them to support this claim, and this has been my experience in conversation with libertarians who espouse the free will defence. Although this would be interesting to examine in detail, given the complexity of philosophical positions on how agents value things, it is beyond the scope of this project.

[35] See Isaiah 59:2 and 1 Corinthians 13:12.

God's presence, but also the future possibility of sin is prevented, so that we can stay in the presence of God. There is a contrast between human relations with God at creation and in heaven. Humans had the capacity to fall by sinning in the beginning, but the redemption of humans means that we are not only able to overcome temptation and not sin, but can be secure in our salvation for eternity because we cannot fall. These are fairly standard theological points.[36] What this means is that the sinlessness of the redeemed in heaven (also called the saints) cannot be accidental. The redeemed do not wake each morning wondering if they will, with God's help, manage to steer clear of sin. Instead they wake secure in the knowledge that the possibility of sinning is no longer something that they need to worry about. So it has to be necessarily true of redeemed agents that they will not sin – sinning is impossible for them.

This means that the heavenly state of humans is not merely a return to the state of innocence that they were created in. If redemption to the heavenly state meant a return to the pre-fall state that Adam found himself in then it would still be possible for the redeemed to fall. There is then, a difference between innocent Adam, and the redeemed. There is also a difference between the redeemed and whatever state the angels were in before they fell. We will need to bear these issues in mind when we explore how God can prevent sin in heaven, because however God achieves this feat it will not do to say something like 'God removes temptations'. Although Adam faced temptation in the garden, there was no tempter present to lure the Devil into his fall. However God achieves necessary sinlessness, it must be by a means that was not present to Adam and the Devil. So spotting the difference between these different states will help us to identify what makes the difference in heaven. Whatever it is that makes the difference, our next question will be why, if God can implement this factor in heaven without destroying freedom, God could not do so earlier. If we cannot answer this question then the free will defence will fail. A key plank to the free will defence is that God could not have done anything to prevent creatures from sinning while maintaining their freedom. So there must be some reason that it is only in heaven that this factor is effective, and also some reason why God could not have created this heavenly state at the outset.

Hopefully what sin is, is something we have an intuitive grasp of. A sin is failure to do something that we ought to do, or the doing of something that we ought not to do. If there is some good that can be done, then failure to do that good thing (unless doing so would prevent some greater good being done) is a morally culpable failure, and a sin. If there is some action that is bad, then unless

[36] For a dissenting approach see Donnelly (1985) who attempts to solve the problem by denying heavenly impeccability.

there is a justification for that badness (such as it being necessary for the bringing about of a greater good), the performance of an action that brings about the badness, is itself bad, and a sin. It is a staple of Christian ethics, that sinning is not only a matter of what agents instantiate in the external world through their actions, but it is also a matter of what they exemplify internally. The familiar example is that murder is a harm, and is bad, and a sin; but also hating somebody, even though we might be horrified by the thought of acting on that hatred or making it known to anyone, is also a harm (plausibly to the possibility of goods being instantiated in our relationships), and bad, and thus a sin. So, considering the heavenly situation, it will not do for heavenly agents to harbour thoughts of misdeed, but to be restrained by God's continued reminders not to do what they are thinking about. The external harm of the sin would be prevented, but there would still be sin in heaven, and thus separation. Instead God needs to prevent sin in a way that means that heavenly agents do not even possess intentions to sin, imaginations of sinning, or imperfections in their internal attitudes towards themselves, others and God. It is worth noting though, that agents can possess internal states required for being tempted without sinning. We do not need to analyse why here, suffice it to note that Christ was impeccable (without any sin) and yet was tempted[37]. So whatever considerations are made of what God needs to prevent in order to prevent heavenly sin, we do not need to add the condition that God makes it impossible for people to be tempted to sin. This is not to say that God could not, perhaps, do this though. Perhaps God can prevent heavenly agents from being tempted, and this helps with securing heavenly sinlessness, but it would not be required unless it was the only way of preventing sin.

Another peculiarity of Christian ethics is the issue of sins of ignorance. In the Old Testament followers of God were required to offer a sacrifice for sins they may have committed unknowingly.[38] This means that the sorts of things that can count as a sin might be more than we ordinarily include with our intuitions. We cannot get agents off the hook by claiming that, or making it the case that, they did not possess the knowledge requisite for performing the actions they should have. Now, it may be possible that sins of ignorance apply when agents should have, and could have, gained the requisite knowledge – a form of epistemic negligence. This would bring them more in line with our modern western intuitions. What this means for heavenly agents is that they will need to not only do the right things, and think/feel the right things, but will also need to know the right things.

[37] See Crisp (2004) and (2007b) for a brief discussion of the issues.
[38] See Leviticus 4:2.

Outline of our strategy

We are in a position to now outline what we will need to do to demonstrate the possibility of coherence between i) and ii). We will need to answer the following questions:

1. What does it take for an agent to act freely (along the lines of creaturely aseity)?
2. How could God prevent sin in heaven without destroying this freedom?
3. Why could God not achieve this in the beginning?
4. Does this leave heavenly agents with significant freedom?

So we will begin by exploring what an action is, and then asking what it takes for an action to be free. We will consider quite a basic conception of action, since we have not developed any detailed thesis about the connection between actions and sin (or responsibility), for example are actions, or sins, always intentional.[39] So our account of action will leave many interesting issues to one side.[40] Then we will consider what further conditions agents need to be under in order to act freely. We will then be able to stipulate that all of these conditions hold in heaven (for the free will defence) and ask what options are open to God to secure the sinlessness of agents under these conditions. As we consider these options, in order to preserve faith in the free will defence we will need to explain why God could not utilise these options before the fall. Our other constraint on these options will be that we require that divine interactions leave heavenly agents with significant freedom. The combination of these two issues can be framed as a question of whether we need to be free to sin, in order to be significantly free.[41] If all of these constraints are applied, and we are left with a possible option for God's sin preventing interaction, then we will have established that the claims that God cannot prevent sin without destroying free will, and that in heaven God prevents sin without destroying free will, are consistent.

[39] So we will not inherit a Davidsonian conception of action.

[40] This is unfortunate, because in the philosophy of action, and in the metaphysics of freewill, it is decidedly difficult to carve off parts of a problem without committing oneself on other contentious issues. But, to be thorough would require separate volumes on action, and freedom (and probably responsibility as well), before we could even get to our theological puzzle. So, we must content ourselves with trying to be judicious as we choose aspects of action and freedom that are central to explaining how God prevents sin, and try to say enough to be able to say something constructive about this theological puzzle.

[41] See Sennett (1999) and Rasmussen (2013) for a discussion of the problem in these terms.

2. Agents, Actions, and Controlling Sources

Agents, sources and the power to do

In order to understand the metaphysics of possible divine involvement in human agency, we must first understand the conditions under which agents can be said to act, and to act freely. We need to distinguish an agent's actions from other things that involve the agent. Agents possess capacities to perform actions, and this capacity enables them to be more than passengers carried along by the unfolding natural physical drama of the universe. Agents are more than conduits through which history pours into the future. Their involvement is more than instrumental; they exercise capacities of agency in order to move the world, and themselves in it.[42] Actions then, are things that agents *do*, not things that happen to agents, and it is this distinction that proves resistant to clear philosophical analysis.[43] For instance an agent may be stood admiring the sunset at the seaside, when a sudden gust of wind blows them forward, into the water. Their entering the water is something that happens to them. Alternatively the same agent may decide that the water looks inviting (it is a hot day, and the water looks cool and inviting) and spontaneously jump into the water. This time the agent does something; they make something happen that would not otherwise have occurred – the agent acts. But for their intervention in the world by acting, they would have remained where they were. However the agent may have been stood, wary of entering the water, when an unusual neural event in their brain occurred, unbeknownst to them, which subsequently caused an involuntary muscle twitch, causing them to fall into the water. In this case it is not immediately clear whether the agent's falling into the water is something that happened to the agent or not.

Consider the case that looks most clearly like action by the agent; the agent exercised a capacity for choice and decided to enter the water. In our example the agent had reasons to do so, and they acted in a way that responded to those reasons. Their response was sensitive to their apprehension of their relationship to their circumstances. They were aware that they were hot, that the water was

[42] This claim is a little stipulative, and does rather beg the case against compatibilists, but following our discussion of the type of freedom a free will defence requires, it is clear that this is the sense of freedom we need to investigate. Frankfurt (1978) suggests the active nature of behaviour as an important indicator of being an action.

[43] Davidson (1971) limits actions to being something that agents do, adding that it must be intentional under some description. We will not follow Davidson with this further condition. I am not sure if intentional action is a sub-class of action or not, but it is a distinction we will not explore, partly because of space, and partly because it may turn out that some of what agents do that is significant, is not intentional. Wilson & Shpall (2012) note that the connection Davidson proposes enjoys lots of support, but that explaining the connection has proven difficult.

cool, that it would be good for them to cool off, and that jumping into the water would be an effective means to that end. In the case where the agent is blown into the water there is a reason for the occurrence too – the force of the wind on the agent is sufficient to topple them. But in this case there is nothing to do with the agent's apprehension of the gust of wind that is relevant to the outcome. In a way they are sensitive to the reason, in the sense that when pushed by the wind, they reacted by falling over. But such sensitivity is entirely passive, and demonstrates no more interesting a capacity than an inanimate object's sensitivity of moving when pushed. When the agent acts they are not passive, they are active at least inasmuch as that the influences over their jumping involve them in a way that allows us to identify the source of the action as being the agent.

In an instance of action, the agent cannot be only passive with respect to the influences over their jumping into the water. If their role is entirely passive, we can straightforwardly identify the occurrence as something happening to the agent. If agents are active, then at least some contribution to the occurrence of their acting has its source in the agent.[44] We have not elucidated in what ways the agent may exert an active influence, but we have identified a core issue of agency: agents control how they act, and part of what it is for an agent to be in control of how they act is for the agent to be the source of what happens when they act.[45] However, explaining what it is for an agent to be a source is no simple feat. It is tempting to start talking about the sufficient conditions for jumping into the water, and claiming that these conditions must be mental, or neurophysiological, states of the agent. However this cannot be quite right on two counts. First, an agent could possess a state that is a sufficient condition for jumping without exercising the sort of control we are interested in, and second there are plausible examples where it is difficult to decide how to identify the sufficient condition of the agent's act. Regarding the first issue, consider a process whose outcome is controlled by a quantum interference device capable of producing a binary random output. When activated, this device produces either a '1' or '0', and the output causes some observable event, like a blue or red light coming on for the '1' and '0' respectively. Due to the nature of such devices it is in principle impossible to predict what the output will be for a single event of the device's operation. This is because there is no sufficient condition that can be identified as constraining the device to output in a particular way – it is purely random. If we set up and activate the device then whichever colour light comes on, the device will have been the sufficient condition of that light glowing; the device will have been the source,

[44] Which does not necessarily mean an internal prior cause, such as a 'trying', such as is suggested by Hornsby (1980) and Ginet (1990).
[45] For a good introduction to this approach to action, focused on agents as a type of source, see Mele (1992) and (1995).

there is no other possible candidate. There is no prior factor that constrained how the experiment could play out which could be the source of the event of the device outputting the signal that it does. So we can say that the device controls which colour light will come on.

Now suppose that a device of this nature exists in an agent. It need not have been put there by a malignant neurosurgeon, in fact, let us suppose that it is a naturally occurring biological phenomenon, an isolated neurophysiological system whose output sends a signal down a neuron to some part of the brain (the details will not matter). The device activates at regular intervals. In the biological case lights do not come on, but the effects are to determine either that the body remain as it is, or that it jumps forwards three feet. Now consider our agent stood on the beach. If the device activates and causes the agent to jump forwards into the sea then the source of this occurrence is not external to the agent. We can ignore for a moment the question of whether being physically internal to the agent's body is a correct analysis of having its source in the agent – let us suppose for now that it is, since we could construct an alternative case placing the device differently in the agent. What we have done is construct a case where the source of the agent's jumping into the sea is the agent (or within the agent), and yet I think we immediately see that it would be strange for us to claim that agent controls their jumping into the sea. We can imagine such an agent finding themself very surprised to be jumping into the sea.[46] So what has gone wrong in our analysis? In the example as stipulated, that agent is the sufficient condition of the jump into the sea, and hence the source, but it is not the right kind of condition to constitute an exercise of control. There is a disconnection between the agent and their circumstances, and the jump. The factors mentioned before, the coolness of the water, the hot day, the agent's desire to cool off – these do not seem to play a role in what the agent does. The agent could just have well have jumped if they were cold and afraid of the sea. It is perhaps easier to see the problem if we consider the same agent lecturing at the front of a lecture hall. Now a jump would be a bizarre occurrence, disconnected from the agent's activity, and inappropriate as a means to any conceivable end the agent may have while lecturing. The key point is that the context of the situation agents find themselves in should matter. If the agent is to be active by exercising control, then it is not having the source of what they do within them simpliciter that matters, it is their capacity for being the source of

[46] If the device has outputs with equal probabilities then the agent may be rather less surprised, their frequent inexplicable jumping being something they have grown used to. So, we can suppose that the output causing a jump is a very low probability event. I do not mean to imply that agents need to possess any sort of phenomenology of acting, such as Ginet (1990) suggests, or might be implied by Anscombe's (1957) 'knowledge without observation'. On the issue of cognitivism about one's own actions see Setiya (2014).

how they respond to relevant features of the context in which they find themselves.[47]

It is tempting to add that it must be responding *appropriately* to relevant features that is important. But what could be meant by appropriately? If we mean, 'most appropriately', in terms of what is maximally rational, or aligns most strongly with an agents desires, or something like this,[48] then we have reason not to add the 'appropriately' caveat. Agents sometimes do things that we find unexpected because we consider them to have reason to do otherwise, or to lack reason for what they do. For example, the agent who fears the sea seems to have no reason to jump, and we might consider their jumping to be inappropriate. However it is possible that such an agent could spontaneously decide that they have had enough of their fear, and courageously (or foolishly) choose to leap. It is at least plausible that agents can do this, and our own introspective awareness of acratic action does seem to admit of possibilities like this. Agents do sometimes do quite unexpected things, even from their own perspective. In this case though, if the agent makes a prior decision to face their fear, then the jump may have just become appropriate for them, given their altered mental states. In this case they may still be doing what is most appropriate, at least, given the considerations of their own practical rationality. The example we would require is an agent whose spontaneously jumping constitutes (rather than precedes) their sudden avowal to face their fear.

Do agents possess the ability to be a source of what they do, in a way that is sensitive to the context of the situation they find themselves in, where what they do is not necessarily most appropriate (in the sense that they have no countervailing reason to do what they do)? Relatively simple examples will suffice to show this is possible. An agent may be offered a plate of biscuits containing chocolate and raisin varieties. There is no feature of either variety that makes it a more appropriate choice for the agent to pick. The agent must do nothing, take one or both biscuits, or refuse the offer (we will assume that this strange scenario is bare of any other considerations, and all other responses would be disconnected from any consideration the agent could posses of the features of the situation they are in). So the agent picks one, eats it, and enjoys it. If we asked the agent why they did what they did, they could respond by giving a reason for their action that relates to some feature of the circumstances of the scenario, "I decided to try a chocolate one" for instance. The agent is responding to relevant features of the

[47] O'Shaughnessy (2008) suggests that there is a connection between perception and action.
[48] A complex, but less naïve description of what someone might want to claim is: in a way that closely aligns with how those features are most likely to provoke the agent to respond given the nature of the agent, their desires for certain ends, and the possibility of attaining any desires in the circumstances they find themselves in, given long term considerations of the effects of their actions.

situation – biscuits are for eating, I have been made an offer, which provokes me to respond to the content of the offer, etc. The agent is the source of their taking the chocolate biscuit, and it does not happen to them. Moreover, they are a controlling source – the action arises from the agent in a way that engages the agent in the features of the scenario in a relevant way.

If the agent had suddenly thrown themself to the floor then this would not have engaged the situation appropriately. It is not that agents could never do this, but remember we are to imagine a scenario with no other factors pertinent to the possibilities. The agent is not afraid someone will throw something at them, has no mischievous thought of pulling a prank, possesses no subliminal desire to take a closer look at the carpet, or any other possibility we could cook up. If the agent did throw themselves to the floor in the absence of any feature of the situation's being relevant to such a course of action – if there was no connection between such a course of action and any possible desire, motivation, or end the agent could aim at, then such an action may have its source in the agent, but the agent could not be a controlling source. It would be similar to the jump instigated by the quantum interference device. If the agent was an avowed chocaholic then they may have strongly wanted to take the chocolate biscuit, but on a whim may have taken the raisin one to try something different. This action would be surprising, but it does still connect with the features of the situation in a way that is relevant to possible ends the agent can adopt as reasons at which their taking of that biscuit can aim.[49] The action would be appropriate in the sense that it is not irrelevant or disconnected from the features of the scenario, but it would not be the 'most appropriate'. So we can dispense with any appropriateness condition in this sense, and explore what it is to be relevant, which seems to capture a condition required for an agent to be a controlling source.

It may seem unnecessary to spend so long considering this appropriateness condition, but it does get us to the point of being able to note an important claim. We should be wary of any attempt to analyse what it is for an agent to be a controlling source that reduces control to terms of normative rationality.[50] Agents, it seems, can do things which are irrational in the sense that they are acratic and fail to promote ends that the agent cares about most, or are actions done for reasons that the agent cares very little about compared to alternatives.[51] For example, an agent may succumb to a relatively weak temptation, even though they set out to avoid it. But this does not mean that an agent is thereby not

[49] I do not want to imply that agents only exercise this sort of control in Buridan's Ass type situations, but we will discuss the scope of agential control later.
[50] Even if that normativity has a subjective basis.
[51] For some discussion of what this possibility entails see Davidson (1969), Sandis (2008), Holton (1999), and Arpaly (2000).

exercising an agential capacity to control what they do. Although acratic action may be an instance of the loss, or rather diminishing, of an agent's rational control, it is no loss of their agential control. Agents can do things that are quite out of keeping with their character, with what we expect of them, and they expect of themselves. But at no point does the distance between what is expected, and what they actually do become so great that we deem them to have lost agential control, for this measure, though it may tell us something about whether a person is virtuous, stable, rational etc., tells us nothing about whether they are exercising agential control and thus performing an action. It is not that we are not interested in these other issues, but they are a matter of *how* an agent exercises their agential control in action, not whether they are performing an action. Care is needed in considering examples though, for it is possible that something unexpected happens because there has been a loss of agential control – perhaps, after all, a malignant neurosurgeon has fitted a quantum interference device to an agent.

This does not mean that agents can do anything as an action, and that control is entirely unfettered, allowing the agent to do (or try) anything at all. There are obvious metaphysical limitations on possible actions of agents. An agent could not fly out of the window and up into the sky. Human beings do not possess this capacity. However an agent could get up on the windowsill and try.[52] Their attempt would not last long, and would be highly irregular, perhaps indicating that the agent is confused about practical, rational or prudential concerns. But it would not make what they do any less of an instance of controlled agency. However there is a scenario where we can imagine an agent who is not suffering any confusion, for whom, given the content of their beliefs, desires, goals,[53] and any other candidates for what constitutes the possible connections between the agent's capacities, self and the world, trying to fly is just metaphysically impossible. Such an agent does not have the ingredients to produce an action like this. Since there is no relevant connection between them, the world, and trying to fly out of this window, the raw materials are not present for the agent to be a controlling source of that action. So if this was to happen, the required ingredient must come from somewhere else, and the source would not be the agent; it would be something that happens to the agent, and not something they do.

Agents, circumstances, and sensitivity

We have used this idea of a relevant connection but have not really identified what it consists in yet. What is it for there to be a relevant connection between an

[52] Trying would be a slightly different action, but closely related as it would aim at, arguably, the same end (see Hornsby for further discussion).
[53] Or even their physical realisers, if there be such a thing.

agent and the world? To move forward we need to know what this relation is between, and when the conditions for this relation are satisfied. In some ways it would be good to be a little vague here, as we might be able to understand enough of this condition without having to commit ourselves on some contentious issues. For example, if our analysis requires us to claim that it is what, within an agent, is required to enable the production of an action that we are interested in, then it would be helpful not to have to worry about whether should talk of cognitive content, or it's physical state realisers, or non-physical (perhaps emergent) states. I think we can be clear enough on the conditions for agency, without having to worry about every detail of the mechanism of agency, so we need not identify the relata in these terms.[54]

When a relevant connection obtains, that connection explains what it is about the agent that enables them to produce that particular action in a controlled way. Whatever the mechanism of action production is, there must be something the agent possesses which provides the impetus for the agent to perform that action. It is customary to talk of desires as the motivators of action, or we could choose to talk of what an agent values, or what an agent perceives, or reasons, as something that resembles their conception of the good (or a good), or of a pro-attitude.[55] But we need not quibble at this point about what it is that might supply motivation. We can use 'desire' as a placeholder term, but are not committing ourselves to any theory of action motivation. Just as a mass does not move without being acted on by a force, so an agent does not act (or try to act, if you do not think trying is an action) without the presence of some motivation that is able to engage with some connected course of action. I add the last caveat because it is plain that an agent's having a desire to drink would not provide any impetus for them to jump on the spot. It is difficult to see how to argue that there must be some motivation with which an agent acts, but I hope it is a plausible enough idea, whether it be a desire proper, or some other feature of agents. It is certainly prevalent enough in our folk psychology, where we often talk about the reason why, or for which, an agent

[54] Although some disciplines try to tackle these issues by first understanding the mechanism, and then seeing what that tells us about our capacities, it seems that there are some necessary pre-reflective assumptions about the metaphysics of different capacities that are required for the interpretation of theoretical models to have enough metaphysical content to engage these higher-order questions. This is not much of an argument, but only a small gesture those who worry that something like neurophysiology should have priority. Steward (1997) even suggests there are fundamental metaphysical prohibitions to this form of analysis. For a recent overview of the issues, see Gibb, et al. (2013).

[55] `I am being deliberately broad here, not wanting to commit to an explicitly Humean theory of motivation, perhaps agents can be motivated to act by reasons, ends, or something else. For a discussion of the Humean theory of motivation see Smith (1987), Schueler (2009) & (2003) and Sehon (1997). For a discussion of desires, and their role in motivation see Mele (2003), Schueler (1995), and Schroeder (2010).

acts, and we will leave it as an innocent enough assumption, and will leave the burden of proof with any objector.[56]

Desires have objects; they are for something, an end at which they aim. This is how we identify desires, often using the locution 'my desire for E', where E is the end. The ends at which desires aim can be internal to the agent, for example a desire to feel warm, and they can be external, for example a desire for a chocolate biscuit. We sometimes express our desire to act a certain way, for example 'I desire to go swimming'. It may well be that we can redescribe these ends in terms that are internal to the agent, for example 'I desire the experience of enjoying swimming'. This need not concern us, because we are interested in what is common to all these: that they express an identifiable end. Note that there is an end that is valued when an agent sees something as valuable, or that is evaluated as good if a moral (or aesthetic) assessment is made[57]. So we are not losing generality by continuing our discussion in terms of desires. Desires are features of agents, and if the agent is to be in control of their action then the desire that provides the motivation (or reason) to act in a certain way must have an end that the agent takes to be possibly achieved by the performance of that action. This seems to be an obvious point; if it were not the case then we could have an agent's thirst (desire for water) providing the motivation for them to give a lecture. Which, in the absence of a convoluted chain of events involving the lecture causing someone to buy them a drink, would be absurd. Such a disconnection between motivation and action may exist in cases where a psychiatric disorder persists, but this is what we would consider a breakdown of agency and control.

We characterised an action as an agent doing something. This 'something' that the agent does can be described in many ways. For example the 'jump into the water' is something the agent does, but the same action could be re-described by citing the bodily movements involved, that the agent 'rapidly extended their legs'. But in each case there is at least one description that identifies an action by the end at which it can be said to aim. An agent can act in such a way as to pick up a biscuit, to throw a ball, to hurt their enemy, to find the product of two numbers (both as a mental action, and externally, on paper), to try to score a goal, to duck under a sudden obstacle, to recoil from a blow, to verbally express their frustration, and to ignore the telephone.[58] Notice that these are not all what we

[56] Pink (1996) has a lengthy discussion of this issue, however he establishes the point through example and appeal to intuition, which will not add much to our present discussion, so readers who find this folk conception unintuitive may consult his examples.
[57] It may, of course, be the case that the 'making of an assessment' is not something that immediately precedes action or decision, nor is concurrent with it. Such assessment may be long standing states the agent has established, conscious or not.
[58] I am not going to provide an argument that this is so, as it is a fairly orthodox proposal, see Davidson (1963) and Pink (1996).

might call intentional actions, and exhibit differing degrees of engagement with an agents rational faculties, some we can imagine as following careful deliberation, others are what we might think of as instinctive. These ends are states of affairs, features of the world (which can include features of the agent), which the performance of the action makes obtain.

Remember that we said that relevance was a matter of connecting agents and the world in the right way to make being a controlling source possible. The end at which an action aims is a feature of the world, and looks to be the right sort of thing to be one of our relata. However, this would not be quite right, for actions sometimes fail to produce the state of affairs they aim at. In this case the end is not a feature of the world, it does not obtain, and in some cases cannot obtain. This would mean that failed actions did not actually aim at what we thought they did, and this is implausible. Instead we mean that actions aim at something more like a state of affairs that an action can make possible, a possible state of affairs. But it is important that we keep the right sort of possibility in mind here. If we mean that the feature of the world agents must possess a connection with to be a controlling source is what is metaphysically possible, then we limit agents to only be able to control their actions when they act in a way that has at least the possibility of success. But this would be too strong a constraint. Sometimes the reason that agents fail to produce the state of affairs they aimed at with their action is because they make a mistake, or are wrong about what is possible. For example an agent could throw a ball out of the window, and the end at which their action aims is the ball's sailing into the garden. However, if the window is closed, then their action is doomed to fail to achieve its end. Does this mean that we wrongly described the action? Well we could re-describe the action as the agent trying to throw the ball out of the window, and in a way this would be right, it is something the agent tried to do. But what was the end at which the agent's action aimed? Were they aiming to try to throw the ball out of the window, or to actually throw the ball out of the window? Although agents are certainly capable of the former, especially when they are unsure of their ability to succeed, it seems problematic to deny that an agent could accurately report the relevant features of their agential control as aiming to throw the ball out of the window.

So what sort of possibility matters? It seems that it is what is possible in the eyes of the agent that matters. If the agent perceives, intuits or deduces that an end is possible, then they can aim at it by acting. So it must be a state of affairs that the agent actually grasps as possible that is aimed at in controlled action. By grasp I do not mean that they have to be consciously, or reflectively, aware of it. It might be a matter of non-conscious perceptions and habits learned in similar circumstances that provide the agent with the relevant states that would constitute their possession of a non-conscious awareness of what state of affairs is possibly

obtainable. For example drivers of cars seem to have an awareness of whether they can brake in time to avoid an obstacle, or need to swerve instead, though people utilise this capacity for grasping possibility with differing degrees of success of course. Now, if an agent threw a ball with the aim of doing something that they thought was impossible to achieve by that action, like making it rain, then there would be something wrong with the connection between the agent, and the end at which their action aims. An agent may be deluded, and think that they can make it rain by throwing balls. But if they do not, then there is no state of the agent that could possibly be a motivator for that action, whose content had to do with the end 'making it rain'. There is no desire present to engage with the features of what the agent could take to be possible action.

However, it is not only what an agent 'thinks' is possible that matters. For example a blind person may claim that they are going to throw the ball out of a window. But if they are in a room they have not been in before, and have no knowledge of where the windows are, or even if there are any, and possess no reason to think that a window is in a particular direction (like a gust of wind), then it would be strange to say they are in control. They might hope to get the ball through a window, but in the absence of any reason to think there is a window where they aim to throw the ball (no matter how flawed or weak that reason may be) it does not seem possible for the agent to be in possession of the ingredients required to be in control of 'throwing the ball through the window' (note: not trying, hoping, etc.). Our earlier discussion noted that agents should be sensitive to the world around them in order to be controlling sources. So it is not only what an agent thinks (or other states relevant to awareness) about the world that matters. It must be real states of affairs in the world that provide the agent with the reasons, or ingredients, they need to surmise, intuit, deduce, or perceive that there is an end at which they can aim. The agent need not always be right about whether, given the metaphysical possibilities, their action will in fact achieve the end at which it aims, nor need it even be a high probability that the action will achieve its end. But there must be some feature of the world, which supplies a reason for taking the action to be a means to its end.[59] The agent must be sensitive enough to the world around them that they have a reason, a minimal one, no matter how weak, or how much it misses other relevant features a more perceptive agent would notice, which is the reason why their action is a means fitted to an end, given their circumstances. These reasons are about connecting means and

[59] 'Tryings' are a strange group of actions, because we are not necessarily concerned with a need to succeed – we can try all sorts of things quite speculatively. So the features of the world required to provide a reason to try something can be very minimal indeed, and are most likely to involve features of the agent (for example I could not try to throw a ball if I was paralysed).

ends by being aware of features of both, and having these features as practical justification for performance of the action.

So, our condition of there being a relevant connection between an agent and the world in cases where the agent is a controlling source, can be spelled out as follows – the following conditions are all necessary for an agent to be a controlling source (along with the condition that an agent is the source, and it is not something that is happening to the agent):

I. An agent must possess a desire (or other motivation contributing state), for a particular end, E.
II. We must be able to describe the agent's action in terms that identify it as aiming at this same end, E.
III. The agent must be aware of some feature of the world that provides a reason for them to take their action as a means to possibly achieve this end, E.[60]

Do agents need to be conscious of reasons?

Now, all of this talk of means-end rationality, reasons and practical justification, may immediately worry us. Does this mean that agents can only be controlling sources when they are consciously, or deliberatively, exercising their capacity for practical rationality? We have already discussed some cases where we might resist saying this, like sudden braking in a car, so some further explanation is required. Although agents do possess a capacity for conscious reflection on their motivations, aims and action in practical rationality, it is not the presence of this capacity in action production that I want to invoke.[61] An agent may be in possession of the states necessary to satisfy condition III, without their being conscious of there being a reason that the means fits the end. So the sense of 'take' I have in mind in III is not one of conscious reflection, or formation of a practical judgement. It is instead a matter of the agent's faculties operating in a way that means that those systems whose job it is to be sensitive to the world are providing reliable inputs to those systems that orchestrate an agent's responses to the world.

[60] An objector may worry that this is too strong of a condition, and that we have not allowed for situations in which agents are mistaken about their awareness of features of the world (like in cases of optical illusion, and displaced agency awareness experiments). However, if an agent is mistaken about their awareness of every feature, in the case where this condition is not satisfied, then how could they be in control of how they interact with the world? It would be like the blind man with no knowledge of the window. The agent could be a source of their bodily movements, and could be in control of how they move their limbs (proprioception still functioning), but they could not be in control of and action whose end involved their limbs interacting with something in the world.

[61] For a contrary position, though aimed at a particularly rich idea of acting, see Hodgson (2012).

It so happens that most of the time we are interested in cases where agents do deliberate, and are perhaps consciously wrestling about which ends to pursue, especially when we are interested in free action, or morally responsible action. But at the moment we are not limiting ourselves to either of these, only to what it is for an agent to act.

An example may serve us as a good explanation by illustration. Consider examples of blindsight. In cases of blindsight agents are not consciously aware of any visual perception, as far as their conscious awareness goes their visual systems are inoperative – they are blind. However the physiological breakdown in the chain from the eye to conscious awareness has occurred relatively late in the system, that is to say towards the awareness end of the chain. The details of this are not important to us, so we can construct a simplification. We will characterise the system as the eyes, which send a signal down the optic nerve to the visual cortex, where there are a number of non-conscious processes responsible for interpreting this data, and relating it to other features of the agent to facilitate motor control feedback, perceptual error correction feedback, and so on. The agent is then able to access the interpretation and bring their conscious awareness to bear on it, and it is this last step that the agent with blindsight is unable to complete. If such an agent is presented with a task of posting letters through a rectangular slot we see some strange results. If we can secretly change the orientation of the slot, we find that agents possess the capacity to orient the letters so that they will fit through the slot. They do not do this with flawless accuracy, but they do so with significantly greater success than trial and error, or luck, would explain. The explanation is that the agent can initiate an action of putting the letter through the slot, but the details of this, in terms of the bodily movements required, are not all ones that the agent consciously presides over. In a similar way a violinist is not conscious of all of the minute adjustments in their arm, shoulder and hand, which are required to produce a range of tones while bowing their instrument. Many systems whose function is to coordinate the activity of agents given perceptual, proprioceptive, and other feedback, function at a non-conscious level. We can imagine a particularly mean instance of the experiment where we shout at agents to pick up the letter and post it, having previously threatened them with an electric shock for failure. Does such an agent satisfy our conditions for being a controlling source?

I. The agent possesses a desire to 'post the letter though the slot'; after all they do not want to be shocked.
II. The agent's action is their 'posting the letter through the slot', which aims at the same end as the desire.
III. The feature of the world that is the orientation of the slot is something that the agent is aware of, and it is a reason why their non-

conscious system that coordinates motor control with visual perception guides the orientation of their hand to be a means fitted to possibly achieving this end.

So, the agent can be a controlling source of this posting of the letter, they are relevantly sensitive to the world and their place in it. If the agent is the source of what happens, and is not passive, that is, we do not initiate the process by fitting electrodes to their brain, or a quantum interference device, then the posting of the letter is an action the agent performs. It is something they do, and is under their control. There are a range of other interesting agent involving occurrences, which qualify as actions. Consider a reflex reaction like flinching when you see something approaching out of the corner of your eye (perhaps a blind man is throwing balls around again). This involves non-conscious processes, as did blindsight. However, it differs from the blindsight case because there is no conscious deliberation about whether to avoid the ball, whereas there probably was when the blindsight agent decided to avoid the electric shock, in other words, it is clearly not intentional. The reflex action is more innate, and is not something over which we are able to exercise direct conscious control.[62] However there is still an end at which the reflex aims, to avoid whatever is approaching, and there is some motivator in the agent of desiring to avoid being hit (though now our term desire may be inappropriate, but remember it is a placeholder for a range of states). Moreover the agents non-conscious systems involving perception and motor-control are again active, and coordinated in a way that the perception of whatever is moving (the feature of the world) does supply the agent with a reason for what they do. If someone was worried about categorising reflexes of this type as actions, then we could push that intuition by considering cases of overlearned motor control. Some bodily movements, for example those necessary for playing a complex instrument like the piano, can, through sufficient repetition, become memorised as programs in our premotor cortex. So when a musician sees the note on the page, and is engaged in playing the music, they do not consciously decide to play that note, and neither do they need to apply conscious guidance to the production of the note. This case is not directly analogous to that of the reflex, but is closely related. If the reflex is in no sense an action, then it is only a short slide to claim that the playing of the note on the piano is not either – both are similarly controlled, have the agent as the source, and do not involve conscious intention. Yet, the playing of a note does seem to be a clear case of action by the musician.

There are whole ranges of instances of action with a similar structure. There are the many types of reflex reaction, like recoiling from a blow, putting out your arms when you fall, and, plausibly, sneezing when you get dust up your nose. These are all things that agents do, and are controlling sources of as they interact

[62] Though there are some reflex reactions we can resist, even if not initiate or control.

with their environment – they are actions. This could be of importance. Suppose that a, rather simple, theory of responsibility held that we are culpable for any action we perform, in this case we would be culpable for spilling a drink on someone because we flinched after seeing something flying across the room. I am not sure many people would be persuaded that this would be an appropriate ascription of culpability, and there may well be further conditions on moral responsibility. But it is these actions on the edge of agency, where there is a minimal degree of control, that can test other theories in which actions feature. Now, there are some reflex reactions that do not exhibit the qualities required to be actions. Consider the knee-jerk reaction. The agent is sensitive to their environment, after all knees do not jerk upon the sun shining on them, but only in specific circumstances. However there is no end at which the knee jerk aims, and it does not engage any desire for an end that motivates the jerk. A knee jerk is something that happens to the agent when struck on the kneecap, it is not something the agent does, and it is not an action. Very plausibly the agent is not the source of what happens, so this could fail on this count as well. We began this chapter by mentioning the case of a sudden muscle twitch caused by an unpredictable neural event in an agent, causing them to fall into the water. In such a case the twitch does not aim at entering the water, and neither does the neural event aim at any end. It should now be clear that the case is not one in which the agent acts, in fact it would be a situation in which an agent has lost control, analogous to someone's suffering an epileptic seizure.

There are some things that agents do seem clearly to be the source of, which fail for similar reasons. For example an agent may nervously tap their foot while working. This does not seem to be something that happens to the agent in the same way that a knee-jerk does. It is a bodily movement generated by the agent. However, it is not something that aims at any identifiable end, nor is it motivated by a desire for any end.[63] Tapping a foot along to a piece of music may well be different, for the agent may be expressing their enjoyment, or helping themselves to concentrate on the music. It is difficult in some cases to decide whether the conditions are satisfied or not, and it may be that cognitive scientists can help the armchair philosopher by identifying candidates for systems or states that can satisfy these roles. But, whatever the difficulty in identifying actions, in short, there are some things that agents are the source of, which are not actions. Therefore a theory of action that relies only on sourcehood simpliciter, without these additional conditions, will mis-ascribe agency in some cases. So as we consider in what ways human agency can be interfered with, or influenced, we

[63] I have heard it said that such foot-tappings, and leg-jigglings, might be non-conscious expressions of some state of the agent. I am not sure about this, but even if it is possible for this to occur, I ask the reader to imagine a case that is lacking this further feature.

need to consider whether these conditions are maintained. Remember that these are not conditions for free action, intentional action, or any other sub-classes of action. We are interested in these other aspect of agency, but the most basic aspect of agency is the power to do things – to act. So we must be clear on this at the outset.

Controlling sources in paradigm cases of action

We should say something about how these conditions relate to cases in which the agent is more clearly performing an action, like when an agent intentionally pushes a button to detonate an explosive, kicks a ball towards a goal, decides to marry their girlfriend, or prays for God to do something. These are examples of actions that we might describe as significant, to varying degrees. Although it is easier to see how the conditions can be met in such cases, it is important that we are clear on how they are, since it is significant actions with which we are primarily concerned. If heavenly freedom consisted in the ability to sneeze and flinch (supposing, rather implausibly, that such actions turned out to be free) then it would be difficult to see how that would constitute a great good, sufficient to justify the evil in the universe. Consider the agent who pushes a button to detonate an explosive. This could be a case where a building is being demolished. The agent desires to press the button, after all it is their job to demolish this building by detonating the explosive, and we have already described the action as aiming at the same end. As long as the agent is aware of some feature of the world giving them reason to take their pushing of the button as a likely means to succeed, for example, they ran the fuses and set up the firing apparatus, then this is a fairly uncontentious case of action. But it is illuminating to consider features that might make us think twice about whether the agent could be a controlling source. If the agent was a passer by, who intensely disliked people who live in the building and wished it would blow up, and happened to see a button nearby which they pressed with the aim of so doing, then they would fail to possess any justification for believing that their pressing the button would achieve that end. They can be in control of pressing the button in the hope that it will blow up the building, but not of 'pressing the button to detonate the explosive'.

This might appear obvious, but it has application: what would we say if it just so happened that the building had been rigged with explosives for demolition, and the workers were showing residents round for a last goodbye (health and safety is fairly lax at the site) before clearing the building to demolish it. We have mentioned a simple theory of responsibility where agents are culpable for actions they perform. This agent did not perform the action of blowing up the building, so would not be culpable of wilfully blowing up people they hate. Not because

they lacked the intention – this they possessed, although were deluded about their circumstances. They fail because they were not in control, and did not perform the action. Some theorists focus on intention in culpability ascription, and the agent may well be blamed for having an 'evil thought/intention', however they did not act on this thought, because this was not the action they performed. What we should make of this is best left to a discussion of responsibility, which we will not begin here, but it does highlight the importance of clarity on these points.

In an alternative case the agent may have been a local official, tasked with pressing the button at a public demolition ceremony. However, there may be reasons that the agent is against the demolition of the building. They may consider it to have architectural value, or have a sentimental attachment to it, and dislike the proposed development that will replace it. Nevertheless, if they have decided to press the button because, having examined themselves, and wrestled with what to do, they feel that they have some small sense of duty to their office, then the desire that aims at the relevant end is present. I mention this case because it demonstrates a case where we might plausibly think that any state that could motivate the agent to perform the action would be swamped by the motivation not to (we have to imagine that the agent *really* liked the building). Although we might judge the agent to be acting against their interests, the agent does not thereby lose their ability to be a controlling source. The experience of the agent, who may hate what they do in this case, is not positive at all. And the placeholder term desire may seem out of place, because almost every fibre of their being cries out 'no' – but note it cannot be every part of the agent, some part must provide the agent with a reason to press the button. The case is interesting because the agent would be considered to be acting out of character, and we might not say that their action reflects them. Sometimes philosophers talk about actions, or free actions, having a condition of expressing the character of the agent. So this demonstrates that care will be needed on this point. Consider an agent who succumbs to the weak temptation to press the button out of curiosity about how big the bang will be. They would more clearly be acting out of character. But remember that we can distinguish agential control, from normative rational control. Although the action does not express the character of the agent, or is not expressive of the agent all things considered, it is expressive of some part of the agent. If an event involving the agent expressed no part of the agent's character (taken to encompass his virtues and vices, conceptions of the good, ends he values, reasons etc.) then the agent could not have been a source, and hence could not have acted.

The case of an agent scoring a goal is often taken to present a difficulty when we consider the control that agents can exhibit, thanks to Austin's excellent paper, *Ifs and cans* (1956) on what an agent 'can' do. I have deliberately avoided this locution. If an agent kicks a ball towards a goal then kicking the ball towards the

goal is an end at which their action can aim, an end that the agent desires, and one that the agent can have good reason to consider achievable by their bodily movements given their sensitivity to their circumstances. But what about the end of scoring a goal, is scoring the goal something that the agent has control over, is it something they can do? Considering control in terms of whether the agent *can* bring about a certain end becomes confusing. After all, the goalkeeper may stop the ball, or the agent may just miss since they are fallible in their exercise of the skill of ball kicking. However, given our criteria scoring a goal is a desirable end at which the agent's action aims, and one that the agents sensitivity to features of the ball, his body, to goal, etc. provides them with a reason to take what they do as a means that will possibly achieve this end. So the agent's action of scoring a goal fits the criteria. This is because we have not included anywhere a criterion of success. Although there is a sense in which we are interested in whether agents have the control to successfully implement a plan, such a scoring a goal (or holing a putt), this sort of control considers the skill of an agent, and the probabilities of various interactions with the world being successful given the skills, abilities and capacities of the agent. But the level of skill-control an agent possesses is not the same as whether an agent is exercising agential control – and the two need to be kept distinct. Agents are not always successful in completion of their actions because they are not the only influence over the way the world evolves.

Some have sought to analyse this problem in terms of agents being able to always perform basic actions, ones that are immune from defeasibility, or are internal to the agent, such as mental actions, or basic bodily movements.[64] I am not going to endorse this distinction, but neither am I going to deny that it may be a useful one.[65] For our purposes it is enough to understand when agents act, and act freely, without having criteria for when agents are in control of the obtaining of certain ends, that is, how much control agents exhibit not over their actions, but over the world. To see why consider a case where a heavenly agent acts with the aim of successfully bringing about some sinful state of affairs. Were God, angels or anything else to intervene (consciously, or even accidentally), then there would still remain a description of the action under which the agent would be sinful – that of attempting to bring about such a state of affairs, or that of performing some action with the intention of bringing about sin. If God is going to secure heavenly sinlessness, he cannot do so by saving metaphorical attempts at the goal of sin.

We should also note that some actions do not affect the world external to the agent, but affect that part of the world that constitutes the agent. We do not need a taxonomy of mental actions, but ends such as finding the product of two

[64] See Hornsby (2004) for a discussion of bodily movements especially.
[65] Basic actions, as a primitive beginning to what agents do, is introduced by Danto (1965), and developed by (among others) Stoutland (1968)

numbers, forming an intention to telephone friend (which seems to be a persisting state of an agent), trying to remember what happened yesterday, and so on, are typical of the sorts of ends agents can do something about by acting. Agents do not aim at the movement of their bodies in order to perform these actions, though (unless one is a Cartesian dualist of some sort) the performance of such mental actions will have an effect on the bodies of agents, in the fine details of their neurophysiology. Agents do not try to change their physical state, or aim at producing these changes in their action. Agents do not have the requisite faculties for sensitivity to their brain states that they would need to be aware of features of the world necessary for having a reason to take any attempted action as likely to produce the desired neurophysiological effect. So such ends cannot be ones agents aim at in action, as they fail criterion III. This points to a non-reducibility of the description of mental actions, to neurophysiological processes. It may be that the same is true of other actions, and perhaps the intentionality of throwing a ball is likewise non-reducible. There is also a difficult question regarding which ends agents can aim at concerning their own bodies. Plausibly an agent can aim at contracting some of their muscles, but cannot aim at the firing of a motor neuron. However such question, while noteworthy, and tangential to our concerns, and thankfully we do not need to worry about such distinctions.

One further concern about our criteria is that agents sometimes report that they do something 'for no reason', which might seem to be a problem as prima facie this could be taken as reporting that there is no end the agent desires, and thus no reason which could possibly be provided as to why they take their action as a practically justified means. Pink (1996) considers the case of somebody who while out walking, sits on a park bench, enjoying a rest, who then spontaneously gets up and continues walking. Is their getting up from the bench an action? Intuitively it seems like we should say that this is an action the agent performs. First of all we should be careful about how we understand the claim of doing something for no reason.[66] Often agents report this and mean by it that they are not performing an action as a means towards some further end – in other words the action is not part of a plan.[67] For example, I may throw a stone into a stream, and when asked why, say that I was trying to hit a stump of wood protruding from the water. In this case my throwing of the stone is something that is part of a plan to amuse myself by attempting to hit the stump. However I could have said for no reason, meaning that there was no objective I hoped to achieve by throwing the stones, I was just doing it absentmindedly, perhaps while thinking about something entirely unrelated.

[66] See also Hursthouse (1991) on how reasons can play a role in these types of action.
[67] See Bratman (1987) for more on the role of plans in practical reasoning and guiding of agency.

In the case of getting up from the bench, that action is spontaneous, and we can imagine that the agent could just as well have risen five minutes earlier or later. Furthermore that doesn't seem to be anything that explains why the agent rose when they did, since nothing significant had changed (we can stipulate that this is the case).[68] The agent did not suddenly notice the time, or feel rain starting to fall. But it is still the case that the agent did have an end at which they aimed – they aimed at standing up. Some actions are reflexive, and are best described as aiming at their own performance. Such actions may be rare, and relatively uninteresting, but are an important class of actions as they characterise an agent's capacity for spontaneity. The agent who rises from the bench in order to stand up can also possess a reason, or motivation for standing up (the same end). They may have suddenly felt like doing so, or have impulsively decided that it would be a good idea to do so although not for some further end, like stretching their legs. Furthermore the agent is aware of the relevant features of the world, the position of their body, that they are sat down, the bodily movements required to stand up, and so on, and these features give the agent a reason to do what they do in order to achieve their end of standing up. The description of the case demonstrates just how minimal the conditions on performance of an action are. It is a reminder that, although we are often interested in examples of action that have other features, we have only analysed what it is for an agent to act. We have not said anything about the varieties of action, intentional, premeditated, etc., nor about the means by which agents may guide their agency and self, such as decisions, plans, practical reasoning, intentions, deliberation and so on. If an agent really did stand up, but had no part of their psychology engaged in awareness of their sitting on a bench, nor of any feeling, desire, reason, decision, inclination, or any other possible motivational state related to standing up, then standing up would not have been an action the agent performed. The agent's faculties that are pertinent to controlling their agency are absent the description, and even if the source of the standing were within the agent, it would not qualify as an action. It would be something that happens to the agent, like the twitch caused by a neural anomaly discussed earlier.

Agents, sourcehood and sufficient conditions

We began our discussion by noting a distinction between an agent being active as a source, and being passive, and only instrumental in the production of some effect or event involving them. We rejected the idea that we could just identify when the agent was the source by identifying when the agent was the

[68] This raises the issue of action explanation, one that we will examine in detail later, but for now we are just trying to characterise the scenario.

sufficient condition simpliciter because it is not only being a source that is relevant to action, it is being a controlling source, and we have discussed, at length, what this consists in. However we also noted that there are cases where identifying the sufficient condition is difficult, and we must now expand on this. Consider a case where an agent has a firmly established habit, such as stepping back from the edge of the road when hearing a car approaching.[69] Our agent has developed an ingrained reaction to these circumstances and 'automatically' steps back when this happens.[70] Given this description of the situation, it is plausibly correct to say that the approach of a car is the sufficient condition for the agent's action of stepping back. Nothing else needed to be added to the situation for the stepping back to occur. We can think of the sufficient condition as the trigger that eventuated what followed, in the case the approach of the car. Contrariwise, if the agent had not been in the habit of doing this, perhaps having never seen a car before, and they consciously deliberated about what to do, eventually deciding to step back, then it may be more appropriate to say that the agent was the trigger for what occurred, it is more clear that the agent settles what happens.[71]

It appears that if sourcehood has to do with sufficient conditions, then in such a case the agent is not the source. The agent is still relevant to what happens, because if the agent had not been structured in a particular way then they would not have reacted as they did. Of course, in some situations we would say that had the agent not chosen as they did then what happened would have been different. When agents choose, in the sense that they exercise something like a two-way power, then the issue of identifying the source is intuitively easier to determine, a choice is in this sense a paradigm trigger. If choice is a two-way power then there are no sufficient conditions for either option (supposing a choice to be between two options), only for the possibility of the agent exercising a choice between them. We will discuss this in detail later when we consider free actions, however in the habit scenario, the agent does not choose, so these types of scenario constitute the more difficult case that might be thought a counterexample to

[69] It is an interesting to consider which actions can be habitual, for example there may be very complex actions that are habitual in sport. In cases of psychiatric disorder such as kleptomania, there are interesting questions about the control agents have and whether they are caused to perform certain acts by events that happen to the agent. However a review of how control is affected by cases of disorder, how actions become habitual, and the mechanisms of habitual action is beyond the scope of this discussion. I do not believe that any of these cases present counter-examples to the material points of our discussion, but I will not argue that here.

[70] Pollard (2003, 2006 & 2010) discusses habits in the philosophy of action, noting that they do not fit well under a Davidsonian reason based scheme. We do not have space to consider the relationship between habits and reasons here, but our discussion may point the way forward for developing this.

[71] See Steward (2012) for a full discussion of the idea of agency as the power to settle what happens. I will not relate and critique her position, but settling is a useful an intuitive notion we can make use of nonetheless.

sourcehood being a condition of agency. Now, although the agent is not the trigger for the event of their stepping back, they are still relevant to what happens. Although the agent is not the source (in the trigger sense) of it being the case *that* the event of their stepping back occurred, they are the source of *how* they respond, and *what* occurs.[72] The car's approaching is external to the agent, and is the trigger of the event of the action's occurring, but what action occurs, that is, how the agent responds sensitively to their environment, is determined by the agent. The agent settles what happens, and it is not settled for them. What we are distinguishing is the sufficient condition for the event of the agent so acting, and the sufficient condition for the agent's action.[73] When an agent is active, rather than passive, and performs an action, then the agent may not be the source of 'the event of an action being performed', but they are the source of 'which action they perform'.

In the case of stepping back from the kerb the agent is active rather than passive because is it something about the agent that makes a difference to what occurs. But it is not making any old difference that will do here. Think back to our earlier example of the gust of wind blowing an agent into the sea. We noted that in this case the agent was passive regarding their falling into the sea – they were not a source of their falling into the sea. It is tempting to say that there was nothing about the agent that made a difference to what happened. But this is not true. In a counterfactual sense, if the agent had a different mass or shape, then the outcome could have been different. We could try noting that there was nothing that the agent could do (as an exercise of agency) to make a difference. But this would also be false as the agent could have knelt down or otherwise altered their surface area, even in complete ignorance of the threat of the wind interacting with them, and could have unknowingly prevented what happened. The agent could do these things and remain passive regarding their falling, or not, into the sea. The difference between the two cases can be drawn out be expanding on the observation that it is how the agent is structured that makes a difference to what occurs. In the case where the agent falls into the sea it is structural elements of the agent like mass, shape and so on that make a difference. However in the case of stepping back from the kerb it is those structural details of how the agent is constituted as a controller, their capacity to be sensitive to relevant features of their environment, to possess ends, motivation, and the capacity to direct matter in relation to those ends. So, if the source of *how* the agent responds is the agent's faculties that constitute their being able to control their responses, then the agent

[72] Steward (1997) has some helpful observations on the pitfalls here.
[73] See Sandis 2012a and 2012b for more on the importance of this distinction.

is the source in the relevant sense.[74] Furthermore, if they exercise those faculties, or capacities, in the right way, as we have discussed, then what the agent does is an action.

Agents, manipulation and failure of agency

We can clarify these conditions by considering some cases in which an agent-involving event occurs, but where it is unclear whether an action is not performed because of some interaction, intervention or manipulation. This is especially pertinent to this project, as we are interested in cases where agency is destroyed, though when 'free' agency is destroyed will have to await further discussion. As well as intervention by other agents, we should also note cases where such interaction is non-personal, and involves only natural physical processes.[75]

The most obvious case of a loss of agency is the remote-controlled agent. If we imagine a series of neural implants that allow us to hijack the agent's motor control systems, then the agent would become effectively a puppet on a string. By hijacking only the motor control systems we can imagine, however implausible the reality of it may seem, that the agent is aware of what is happening, and may well look on what their body is doing in horror. They are disengaged from the production of the action, probably horrified at their inability to regain control. The sourcehood condition is clearly broken; it is a situation that is effectively a sophisticated case of me moving the agent's arm by tying a string to it and then pulling on the string. In a slightly different case we could imagine some counterfactually sensitive neural devices, such that the hijacker controls the agent's body, but should the agent try to regain control, or exert themselves to stop or change the action, the devices will terminate their influence and the agent will be able to do as they please. This is a twist on well-known counterfactual intervener cases. Now, should the agent remain in the control of the hijacker because they omit to exert themselves in regaining control, then this would not make the

[74] Note that this is a condition on action, not free action. There are some cases where libertarians will be concerned that the agent may not be ultimately responsible for the way they are constituted to respond, such as responses that are genetically hardwired like the response of a newborn to 'walk' when held standing on a surface (although newborns may not be the best place to look for cases of agency). These are cases like the habit case, but where the agent never exercised any choice about the development of the habit. In due course we will discuss whether agents who are hardwired to respond in a certain way could be considered freely acting. For now, since it is a point of likely confusion, it is worth pointing out that even if they are not, this does not mean that they are not still acting, and on this account, they could be. Therefore nothing we have said so far commits us to any form of sourcehood incompatibilism, in contrast with other accounts of action that also focus on the power to settle, such as Steward's (2012).

[75] Especially as some have charged various positions in the free will literature of unduly anthropomorphizing natural influences over agents.

movement of their body any more of an action, it would still fail for the same reasons as before. Even if the agent was not observing in horror, but was pleased to see something being done which they had never had the courage to do, the occurrence would remain a non-action for the agent. The closest a natural phenomenon may come to breaking the criteria for action in a similar way is something like electrocution. The source is external to the agent, and causes muscles to contract, and the agent's body to move, without engaging any of the agent's faculties for control, often bypassing the central nervous system and directly causing muscle contraction by exciting proximate motor neurons.[76]

Another case where control is influenced is coercion, such as when an agent has a gun held to their head, and is told to do something, such as handing over their money.[77] Now prior to being threatened we can assume that the agent had no desire to hand over their money, but agents need a desire for the end at which their action aims in order to perform an action.[78] However, once threatened the agent soon finds that they have reason to hand over the money, and a strong desire to do so. It may well be a desire they wish they did not have – to be more precise they are probably glad that they have the desire, otherwise they would be unmoved by the threat, and probably shot, what we mean is that they wish they were not put in a situation of having to act on a desire like this. So agents under coercion fulfil all the criteria for action, they possess a desire to hand over the money, which is an apt description of the end at which their action aims, and an end which engages their sensitivity to their environment. Some analyses of the agent's motivational state may worry that the agent's desire is really that they not be shot. However we are not interested in which desire is strongest, or the agent cares about most. It may be true that the agent desires to avoid being shot, but this entails a further desire to hand over the money, one that the agent would be impotent to secure their safety without.

There are natural circumstances that agents may find themselves in that are analogous to coercion. For example a volcano may be about to erupt on the agent, 'forcing' them to run away. The agent is, of course, not really forced (as the sad cases of the incautious and overly inquisitive attest to), but is provided with strong reasons, and strong motivation, to run away. If the agent acts in the gun case, there does not seem to be any relevant difference to the volcano case, therefore they act there too. There are less extreme cases, such as a mother who discovers

[76] Readers familiar with alien-hand syndrome may want to think about this as an interesting special case. Although the agent's consciousness is disengaged with the movements of the alien hand, it is still the case that these movements are controlled, and have the agent as their source. So alien hand movements may indeed be actions, though perhaps not free ones, nor ones for which agents are responsible.
[77] Kane (1998: 64) calls this constraining control.
[78] Or other motivating state, desire is still a placeholder term.

that her children are trapped in the burning building, and runs in to save them. Here the desire to run into the building is something that the mother, as is normally supposed in such examples, is unable to escape due to the strength of her maternal instinct. But no matter the reason for her desire, the key is that she possesses it, and acts towards the same end in a way that satisfies the criteria. It is worth noting that we can continue to cite examples, each less extreme than the last, from Luther's "here I stand, I can do no other" all the way to the addict, and even beyond to the person with a penchant for chocolate cake. It is equally true to say of all of them, that they are performers of actions, as long as each case does not fall foul of the criteria through any strange details like suffering from an abnormal delusion. We should also note that the influence over the agent need not be as sudden, or proximate as pointing a gun, the influences of an agent's motivation and reasons may be slow and gradual. While these issues may have a bearing on questions of the extent of moral culpability, or of whether agents could be reasonable expected to make other choices, they have no bearing on whether the agent is performing an action or not.

Another way an agent could be 'forced' to do something would be through manipulation of their desires.[79] Desires could be implanted through neurophysical manipulation, or by hypnosis. A less extreme example would be the manipulation of an agents desires by behaviour modification through positive reinforcement. In this scenario we have a covert plan to reward an agent with the aim of establishing a desire to perform a certain action towards a given end, or as an end in itself. Now the agent may have some other desire to perform the action at the beginning of this process, because they may aim at a different end. For example an agent may eat healthy food because they will be fined for eating unhealthy options, but through behaviour modification may come to desire the healthy food because they prefer its taste, or find they have a craving for fruit rather than cream buns. In the case of an improved diet we may be untroubled by the prospect of manipulation, but it is not significantly different from cases where agents may be manipulated to perform actions that we might worry about, such as giving money to somebody, or being drawn into a sexual relationship. Suppose an agent has been manipulated to give money to somebody. They do still possess a desire relevant to satisfying the criteria, such as a desire to give money to Bill, and their action aims to secure this same end, and we can assume that the agent is sensitive to Bill and the relevant features of money giving. So even though the agent was manipulated they still perform an action, though, once again, we might raise questions about whether such an action is free.

[79] These cases are similar to Kane's cases of non-constraining, and covert non-constraining control.

It is interesting to now consider a case where the agent is manipulated by a more direct implantation of a desire, such as through hypnotic suggestion. Cases involving hypnosis are complex, as they can involve more than implanting a desire. It may be more correct to say that desires are adjusted rather than implanted, but we could imagine a case involving a malignant neurosurgeon who really can implant desires. In such a case, as long as the agent really does possess the desire, then they would still be acting. I add the caveat 'really does', because there may be cases of manipulation that make it appear that the agent has a desire, when really they only appear to because their behaviour does not stem from any motivational state but is sophisticated neurophysiological manipulation like the remote controlled agent case – so it may be difficult to assess whether agents have the desire. If a desire is suddenly implanted it would need to be accessible to the agent's faculties of control, that is, they would need to be able to have the desire play a role in their practical reasoning, or to aim at the end the desire is for. Such radical manipulation could involve a large-scale reformulation of the agent's states, as many desires could be altered. But even in this case, though the agent may have been changed beyond recognition, they still perform actions. Manipulation calls to mind control by another agent, but again, such changes in agents could be brought about by natural, non-personal means, such as agents who experience a severe alteration of personality, tastes and so on, after experiencing severe brain trauma.

In other cases of manipulation by hypnosis the behaviour of the agent is altered in a different way. For example the agent may believe that they are holding an apple, which is good to eat, rather than a raw onion, which is not. In this case the agent's desires have not been altered, but their ability to be sensitive to their environment has. They lack the ability to perceive their environment rightly, and therefore cannot be aware of a feature of the environment relevant to their possibly achieving the end they aim at. So in this case the agent is moving their body, but not performing an action, at least not the action of eating an apple, or even of satisfying their hunger. Agents suffering from a perceptual breakdown fail to satisfy criterion III. So if it were possible to hypnotise an agent into thinking that they were preparing vegetables, when in fact they were murdering Jones, then being able to determine that the cutting of Jones is not even an action the agent performed would be very useful for determining the ascription of culpability. It would also, presumably, be a comfort to the agent to know that they need not suffer by thinking that they did something terrible, even if manipulated to do so, but instead that the terrible thing was not even something they did.[80] It is not the

[80] This presumes quite a lot about the connection to responsibility, which I do not want to be drawn into – the case is just interesting in this regard, even if someone claimed something different about responsibility.

presence of a personal intervener, the hypnotist, that determines that the agent did not perform an action. There are cases where there is a perceptual breakdown for natural reasons, for example an agent may be suffering from a naturally occurring hallucination, perhaps that Jones is a pumpkin, meaning that there is a breakdown in their ability to be aware of features of their environment that could provide any reasons for taking their action as a possible means to an end. This is interesting because cases of delusion and hallucination do not involve only a breakdown of agents to orient their actions, but furthermore to even perform an action.

Perceptual breakdown cases are not always easy to resolve however. Consider the case of Ken Parks, who one night got up, got into his car, drove across town, stopping for red lights, and avoiding obstacles, entered the house of his relatives and stabbed two people before returning to his car, putting the bloodied knife on the back seat, and driving away. Ken Parks, though, was asleep, and all that occurred was a case of extended global automatism (essentially complex activity while sleep walking).[81] Since the agent was not conscious, we might think that they were not able to be perceptually aware. However, he did manage to stop for red light, avoid obstacles, use doors, pick up a knife by its handle, and so on. So at a non-conscious level he was aware of features of his surroundings. What is less clear is that Ken possessed any desire for anything that would be a candidate for an end at which his 'actions' aimed. Whatever motivational states were possessed by Ken, they did not involve stabbing his relatives, who he liked and got on well with, and it is speculated that he had a non-conscious 'repel-threat' response when interrupted during his sleepwalk, though what got him to the house in the first case is unclear. The point is that above the suspicion that Ken did not perform an action of stabbing, murdering etc., because he lacked the relevant desire, there is also a worry about a breakdown of criterion III again. Although Ken was aware, albeit non-consciously, of some features of his environment, he seemed to be unable to be aware of some of the features that would have been pertinent to an action described with a specific end in mind. For example he may have had a non-conscious motivation to repel an attack, or threat, but he lacked the requisite awareness to provide himself with a reason that he was under attack. The case of Ken Parks is unusually complex, and without a thorough examination of the relevant psychological and neurophysiological issues, we cannot rule definitively on it.

A related example is that of an agent who jumps up and runs across the desert because they think that they see water – but, of course, there is no water, and they have only seen a mirage. In this case the agent has a perceptual experience, which

[81] Levy (2014) has a thorough examination of the Ken Parks case. For those who can't bear the suspense, he was acquitted.

they interpret as the presence of water. They are wrong though, so does this mean that they are not aware of any feature that would give them reason to take their actions as a possible means to fulfil the end, which we can take to be satisfying their thirst – is criterion III broken again? Well, even though the agent is wrong, they are aware of a feature of their environment which is relevant to whether their action would possibly achieve their end. They see the refracted light, and this light gives them reason to investigate. If they saw a heap of dry bones on a dune, and ran towards those with the same end, then we would quickly worry about the agent's sanity. However, even though the agent is wrong about the presence of water, it does seem that they possess a justification for what they do, a justification, or reason, that is grounded in their awareness of the reflection they see. Therefore criterion III is not broken, and the agent does perform an action (which I hope we intuitively thought anyway). The case is important because it highlights that the awareness and reasons agents need to possess to be performing actions are defeasible – agents can be wrong. However in cases where the agent cannot be aware of features of the world, such as when experiencing a hallucination, they are not only wrong about whether Jones is a pumpkin, they are not aware of any feature of the world, they only think that they are. So it is not only a question of whether agents take themselves to be aware of something, but whether there actually exists some connection between the agent's faculties and the way the world actually is. In the case of hallucination, this connection does not obtain, in the case of a mirage it does.

Criterion III could also be broken by a breakdown of the agent's capacity for practical rationality. An agent could have a perspicuous awareness of their environment, but if they lack the ability to relate ends to possible means then they would not be able to take a feature of their environment to provide a reason to take their action as possibly achieving their end. Cases like this seem to be quite rare, especially since agents can be very bad at practical reasoning, exercising very poor judgement about how possible means connect to ends, but are still acting as long as the connection they take themselves to make is grounded in some feature that they are aware of. However, it does seem conceivable that a hypnotist could suggest to an agent that a suitably disconnected (not just poorly connected, for then there could be a reason to so act, albeit a bad one) pair of means and end, are in fact connected, for example that throwing books out of the window is a means to reading them. I do not think that we would consider agent under such conditions to be performing the action of reading a book. If we replace the hypnotist with a naturally occurring neural disorder, the same analysis applies.

Some of the most notorious manipulation cases are Frankfurt style cases, though they are normally examined in terms of whether the agent is free, or could have done otherwise (questions we will turn to in due course). The normal

protagonists are Jones and Black. Black wants to ensure that Jones performs a certain action, for example shooting Smith, but is prepared to see if Jones will do so on his own. If Jones decides to pull the trigger on his own, then Black merely watches events unfolding. However, should Jones decide not to, then Black acts as a counterfactual intervener in some way to ensure that Jones pulls the trigger. There are a multitude of such cases detailing different ways in which Black might be able to assess whether Jones has, or will, decide not to pull the trigger, and also a multitude of ways in which Black can interfere with Jones to ensure the pulling of the trigger. Let us consider a case where Jones is looking down the gun, and has a five second opportunity to fire the gun to kill Smith, who is passing by, so Jones will either decide to fire the gun and pull the trigger, or will omit to do anything and the window of opportunity will pass. Black has implanted a device in Jones' brain so that Black can tell whether Jones is initiating decision-to-fire activity that will cause the pulling of the trigger. If Black does not detect any activity then he will, at the last nanosecond when it is now too late for Jones' initiation of motor activity to cause the trigger to be pulled in time, activate a second implanted device that causes a signal to be sent from the motor cortex to the muscles involved in pulling the trigger, thus securing the shooting of Smith. (In some variants of cases like this Black does not need to activate the device, since it is set to activate at a certain time, or in a certain situation, unless Jones acts first, cases like this will follow a similar analysis.) In a case like this Jones is not the source of the pulling of the trigger since his faculties of control are disengaged from the event of the activation of the second device. Jones' desires do not play a part in bringing about the action, the output of his motor control systems are influenced in a way that effectively rules out their contribution. Therefore in Frankfurt-style cases such as this Jones does not even perform an action when the counterfactual intervener is activated, let alone do so freely or responsibly.

However, not all Frankfurt cases involve the same overruling of the agent's faculties of control. We can imagine a similar scenario where the second device leaves the agent's faculties for control intact, and connected to the production of the eventual pulling of the trigger in the same way they would be if no intervention had occurred. This could be achieved by having the second device give the agent a strong desire to pull the trigger. If this desire was then effective in producing the pulling of the trigger, as Black desires, then it looks like all the criteria are satisfied, and the counterfactually influenced pulling of the trigger is indeed an action (though there may be concerns about whether it is free). The case becomes similar to the ones of desire implantation discussed above. However, the point of Frankfurt-style cases is that it is supposed to be impossible for the agent to perform any alternative action, in our case refraining from pulling the trigger. Now, if Jones was looking down the gun, but had no desire whatsoever to pull the

trigger, then it would be impossible for him to so act. Providing him with the desire required does make it possible for Jones to now perform the action of pulling the trigger, however it does not make it impossible for Jones to refrain from so doing. We have already noted that agents do not need to act on their strongest desires. This is not because of concerns about freedom, but purely to do with what an action is. So a device that provides Jones with a desire to pull the trigger does not necessitate his pulling the trigger, even if it is a strong desire. Perhaps the device could remove all other desires, even the desire to refrain, so that the agent is left with no other choice? I am not sure that such a state is achievable by intervention, but if conceptually coherent, then perhaps now the agent will perform the desired action, and could not do otherwise.

However, the timing of the action matters as well, since there is a window of opportunity. We might wonder whether such desire manipulation is really enough to ensure that Jones does not miss the window of opportunity. It may well be that in order to secure Black's designs for the death of Smith an intervention is required that goes beyond desire implantation alone, and once we start to add other features to the device, such as the initiation of some neural event, there is a threat that the sourcehood condition would be broken again, and that the pulling of the trigger would no longer be an action of Jones. The initiating of a neural event by anything other than the agent's own faculties looks like it threatens sourcehood, so this raises a doubt about whether Frankfurt-style cases really do show that agents can find themselves in a position where they perform an action but could not have done otherwise.[82] Although the agent could not do anything to ensure that the trigger was not pulled, intentionally or not, there is a sense in which the agent could do otherwise – although the trigger is pulled, and Smith is killed, the agent fails to perform the action. Note however that such Frankfurt-style intervention could be used to prevent an agent from performing an action. If desires are manipulated then the intervention can make possible some actions which were not possible before (though not necessitating them), but can also prevent the agent of performing actions with certain ends by removing any motivational basis for an action being performed for that end. Since the issues of timing and necessitation do not occur, the sourcehood of whatever the agent does do are intact.

The last kind of case is one that involves breaking criterion II. For this criterion to be broken, the agent must do something, but we must be unable to describe it in terms that identify it as aiming at an end that the agent has. Cases like this are what we might call accidental actions, though the term action here is misleading. Consider the case where I leave a rake out in my front garden. In the

[82] See Alvarez (2009) for a similar objection.

morning the postman comes, and steps on the rake, injuring himself. The question is whether I performed the action of injuring my postman with a rake. We may suppose that I really dislike my postman, and would like it if he was harmed, in fact I would like nothing better than to see him step on a rake in my garden and injuring himself, though this is a desire that I feel I should not act upon. Now, in our case, I leave the rake out through neglect or absentmindedness. Leaving the rake in the garden is something I do, however injuring my postman is not, because we cannot describe my leaving the rake out in terms that identify it as aiming at this end. The fact that the postman injuring himself matches my desire is a happy accident, at least for me, if not for the postman. Therefore the second of our criteria is broken. I may still be responsible or culpable, for the injury to my postman (intuitively this seems so), which hints that agents can be responsible for things that are not their actions, but we will not explore that issue here.

The stepping on the rake, and injury, is a result of my action, and it is not something I did. If I had run at my postman, rake in hand, and taken a swing at him, then my action could appropriately be described as aiming at the end of injuring the postman with my rake. The distinction between an agent's actions and the later, or further, results of their actions (since results can be occurrent) is an issue that we find in the problem of action individuation.[83] The classic case is the agent who flips the light switch, thus making the bulb come on, illuminating the room, and startling the burglar who was hidden in the room. I do not want to enter into a full discussion of action individuation now, but we can note how our discussion may inform this issue. The agent's action may be described as aiming at flipping the switch, turning on the light, and illuminating the room, since the criteria can be satisfied when this is the end under which the occurrence is described. However the startling of the burglar is not an action the agent performs, since it fails to meet, perhaps all, of the criteria (I hope this is obvious after so much discussion of what it takes to satisfy them, so will not labour the point here). The startling of the burglar is a result of the agent's action then, but is not something the agent does – it is something the agent makes happen as a result of what they do. Given the appropriateness of the agent reporting something like, "I startled a burglar" it is clear that we are not very precise in making this distinction in our everyday dealings with agents and agency. However, it is a useful distinction when we require more precision, for example when difficult issues of ascription of culpability arise.

[83] Ruben (1999) gives a good outline of this problem.

3. Free Actions and Alternative Possibilities

Being free to do what we want

Now that we have clarified what it is for an agent to act, we must now turn our attention to what it is for an agent to act freely.[84] One important sense in which agents may be free is in regards to the question of whether an agent is free to do as they want.[85] Sometimes an agent may not be free because some external constraint prevents them from performing the action they would like to perform. For instance, I may want to go to the theatre to see a play, but find that I cannot because somebody has hidden the tickets. My freedom of action is curtailed, because there is some action that I want to perform, and would perform if circumstances were different, but cannot. This issue is to do with the scope of open possibilities over which my action is free to range, and is often discussed in terms of social and political freedom. There is a wide range of factors that can constrain my freedom of action in this way, and they need not be personal, like another agent hiding my tickets – a gust of wind could have blown the tickets into the fire and destroyed them. It is nearly impossible to avoid both limiting the possible range of free action of others, and having our own possibilities curtailed, because almost anything we do impacts future possibilities for ourselves, and others.[86] On the other hand, our actions do not only curtail possibilities, they also make things possible that would not otherwise have been open to us. There are interesting questions about how we ought to act, about promoting certain possibilities, and not others, but these are questions of ethics and practical rationality, and about what ends are good, valuable and worth pursuing. This sense of freedom is not about what it is for an action of an agent to be free, but rather about which actions and agent could be free to perform – call it freedom of possibilities. There may be many things that I am not free to do, but in a way, the fact that I could try to, or want to do, these things shows that this is not the issue at stake for heavenly freedom. If God 'hid the tickets' every time I was about to do something sinful, it would not eliminate sin, only the ability of agents to succeed in achieving sinful ends. This points us in the direction of the freedom of an agent to choose at which end they will aim to act – what the agent wills.

However there is a related claim, normally the focus of compatibilist accounts of free action, which may be more to the point. We might ask of an agent whether

---

[84] For excellent overviews of the issues see collections by Kane (2011), (2002a) & (2002b), and Watson (2003).

[85] O'Connor (2010) notes that we find influential seeds of this conception in Hume (1748) and Edwards (1754).

[86] Considerations in a similar vein have led some to claim that only a minimal amount may be solely up to an agent, see van Inwagen (1994).

they were free to act in the way they prefer, given how they are. This is a matter of what choices an agent can make in view of any constraints upon them. When an agent is free they should be able to act in a way that accords with their all things considered preferences given their character, desires, reasons and so on. It should be the way the agent is rationally and motivationally constituted that determines what happens, they should be free to act according to their judgements. Thus this sense of freedom has to do with the power of self-determination, that is, that the self is free to determine, or guide, what action is performed, so that action expresses the agent. We can call it freedom of action guidance.[87] However it turns out that this condition is relatively easy to satisfy, and even agents in cases of compulsion still turn out to be free.[88] Consider a case where the agent is being compelled to hand over some money by having a gun held to their head. The agent hands over the money - did the agent act freely? We previously decided that an agent in this situation does perform an action, and this action does accord with how the agent is psychologically constituted at the time. The agent may well not want to hand over the money, but they want even less to be shot, so on balance they make a rational and informed judgement to hand the money over. Since what they do accords with their beliefs, desires, reasons and so on, they are free in the sense that their action is expressive of their all things considered preference – they possess freedom of action guidance.

However, it seems that we are loath to say that an agent who is so compelled is really free, but why? It is clear that the agent is being forced to act in a way that they would rather did not happen – that is, they would rather not even be in the position of being threatened, they would prefer this, and the associated keeping of their money. The agent is being forced (not literally, but is under extreme psychological duress) to pursue an end that they have not chosen. They are being forced to hand over money, and using their money in order to preserve their life under these circumstances is not an end that they value for their money. They would rather spend the money, or put it to some other use, and whatever that use is, it is something that the agent is being forced to give up as an end. By saying they do not value the end they are being forced to act upon we cannot mean that this is an end that they lack any motivational basis for, otherwise it would not be an action. Nor can we mean that they do not possess a strong, or strongest, motivation for this end. If this is what we meant then we would be claiming that agents are free only when they can act on ends, or reasons, for which they have the strongest motivational basis. Besides, with a gun held to their head it certainly seems true to say that the agent probably has their strongest motivations, and reasons, directing them towards handing the money over and saving their life.

[87] This is a gesture, of course, to Fischer's useful terminology of guidance control (1994).
[88] Strawson makes this point (2011).

Instead we mean that there is something about how the agent is constituted that means that the action of handing over money to preserve life is not as good as the end of spending it as they might have otherwise intended (we can suppose that they were going to spend it on a present for a friend, not to buy drugs, or some other questionable end). 'Good' is a notoriously difficult word to define, but for now we needn't concern ourselves with why agents do, or even ought to, conceive of something as good. It is the effect of their so conceiving that is important, if an agent conceives of something as good then they value it differently relative to other things, they might see it as something worth sacrificing other things for in order to pursue, or something to become attached to. It is not a matter of forming a practical or ethical judgement, these seem to be much more circumstantial. But an agent's valuing something as good is a matter of structuring their conception of the world around them, and their relationship to it, and this in turn gives structure to their practical and ethical rationality.

So, under compulsion, the agent is forced to adopt an end that they considered to be bad, or worse, for them. The agent acted when they handed over the money, and in the sense that they acted in accordance with their own preferences, they acted with freedom of action guidance. However they did not act freely in the sense that their choice between ends, regarding how to use the money, was not one that was made on the basis of what they thought the money was good for. Under compulsion the agent was placed in a position where their agency was manipulated in a way that meant that their strongest reasons and motivations were directed at ends that the agent did not value as good, and the ends which the agent did value, and that were expressive of their character, plans, values and intentions, took a back seat. Obviously the agent did value saving their life, and so there was something of value in handing over the money, but handing over money to save their life is not an end that the agent would choose based on their conceiving of it as a good choice of end (or of purpose[89]). Note that it is not that the agent was unable to act in any other way, or was unable to act in a way that upheld their character.[90] The agent is still the source of how they act, and could have done otherwise. The agent could always have acted in such a way that caused them to be shot – there are cases of people doing just this, people refuse to hand over what are sometimes quite small amounts of money to muggers, or bank clerks refuse to hand over money to robbers. Perhaps the agent thought that

[89] 'Purpose' and 'end' are close synonyms in action theory, however sometimes the term 'purpose' has a helpful tendency to focus our attention on an end's being intentionally adopted by an agent, which may not be true of all ends. So I will sometimes use both to try to cover the broad range of types of goal agents may direct themselves towards.

[90] I am using 'character' in a rather loose way here, to gesture to the set of things that encompass an agents natural dispositions, inclinations and so on, we will have more to say about character later.

handing over the money was the wrong thing to do, and that doing the right thing was just worth more than their life. Or perhaps the agent didn't think that the compeller would really pull the trigger. In such cases it is possible that the agent possessed an even stronger motivation to act in a way that eventuated their being shot than they did to give in to the demands of the compeller. So the agent got to do something that, at least in a sense, they preferred, so why do they count as suffering from a freedom diminishing compulsion?

The actions of the compeller made a change to the agent – they changed the way that the agent related to ends that they value, the means apposite to those ends, and the practical and moral rationality of pursuing different ends, and, crucially, how these all relate to the agent's conception of the good. The agent did not exercise any control, any aspect of agency, in the production of this change. Such a change constitutes an alteration in the capacity for self-determination of the agent. It does not alter how the agent is able to exercise self-determination in terms of limiting possible actions (the agent can, after all, refuse and be shot), so they still possess freedom of action guidance. But it does alter it by forcing a change upon the agent regarding their ability to have reasons to pursue certain courses of action based on ends that the agent actually values as good – the agent does not possess freedom of value-alignment.[91] Freedom of value-alignment is the freedom of an agent to align their motivational bases, purposes, ends, and the structure of their practical rationality[92], with their conception of what is good.[93] This condition is supposed to rule out instances where an agent is compelled, or manipulated, to desire, aim at, choose, or adopt an end that their prior constitution as an agent would lead them to judge as bad. By constitution of the agent I mean the sum of their beliefs, desires, valuings, and other attitudes, which structure allows them to make personal judgements about whether something is good.[94] For example the agent having the gun pointed to their head is put in a position of having strong reasons to adopt a purpose which they think is bad – having to hand over money to save your life. An agent who was hypnotised and given the implanted purpose of wanting to murder their spouse (if such a thing is even possible) has a purpose foisted upon them, which they conceive of as bad. An agent like Jones in Frankfurt-style cases may also fail to possess freedom of value-alignment. If the changes Black makes to ensure that Jones pulls the trigger

[91] A similar connection between freedom and value is explored in Fischer (2012).
[92] I do not mean that the agent can change the 'rules' of practical judgment making, but rather those inputs relevant to deliberation and consideration of practical problems. This is rather vague, but an account of practical rationality is beyond the scope of this project. We will have some things to say about it, but must beg the reader's forgiveness for not exploring these issues fully here.
[93] See Bratman (2000) on the connection of valuing and agency.
[94] Not that such a use of good is supposed to be broad, and can include value-judgements such as whether something is beautiful and so on.

to kill Smith are ones that constitute the implanting of something that is at odds with, and perhaps replaces, that part of Jones' psychology which is the basis of his decision not to pull the trigger, then Jones would fail to be free, just as in the hypnotist case. There are some Frankfurt-style cases, questionable in their coherence, in which Jones fails to pull the trigger, but does not decide against it. The Jones in such cases may need a different analysis concerning their freedom.

These examples have characterised the absence of freedom of value-alignment as involving intervention by other agents, and although it is often these cases that interest us, they are not the only ones. It is possible for agents to act in such a way that they diminish their own freedom of value-alignment. Consider an agent who decides to take some highly addictive drug. We can assume, reasonably, that they act with freedom of action guidance in so doing. However the effect of taking the drug, and the ensuing symptoms of addiction, may be ones that the agent finds quite unforeseen. Perhaps the agent was overly confident about their ability to quit whenever they wanted. The effect of these symptoms on the agent's psychology will be ones that constitute a breakdown of the agent's freedom of value-alignment. They might find themselves strongly motivated to do things they are ashamed of, such as the way they treat their family, or stealing. They do not have the ability to align their practical rationality in a way that respects their conception of the good. Of course, over time, such an agent may find that their shame, values, character and conception of the good, such as they were, become eroded. Their freedom of value-alignment returns because the values they want to align their ends with are not ones that their symptoms and circumstances get in the way of.

An agent could find themselves in a similar predicament to our examples through some non-personal event occurring. For example if an agent's house was struck by lightning, starting a fire, and trapping their children inside, then this could place the agent under the sort of unexpected psychological duress that is analogous to having a gun held to your head. We could say that the agent was compelled to try to save their children, perhaps at the cost of sacrificing something of great value. The change in circumstances was beyond the agent's control, and impacted their ability to align their ends with things they valued as good, so they lost freedom of value-alignment. The agent is placed in a position of having strong reason to adopt a purpose which they think is bad, given what they value, and their conception of the good. It might be wondered whether the impact of natural events is really something that it is pertinent to speak about in terms of freedom and compulsion. However, we could quite naturally speak of a case like this in terms of tragedy, which does involve the loss of things that we value, such as opportunities to have and do certain things. But, what about natural events that are not quite so dramatic, and do not involve personal tragedy. Imagine a case

where the agent wants to tend to the roses growing in their garden. This is something the agent values, as they are a collector and cultivator of rare varieties of rose. Now, if there were a sudden rainstorm, then the agent would be unable to go out into their garden, and the roses would be subsequently damaged. Well, actually, this may be overstating the severity of the situation, surely they are not compelled to stay indoors, after all it is only a bit of water. So perhaps we should increase the severity to be a strong gale that is bringing down braches from trees and the like, or something similar with a moderate element of risk. In both cases there is something bad about pursuing their original aim, though to different degrees, and in both cases what the agent has to forego is a loss, again to varying degrees. But the agent doesn't seem to be compelled; wanting to avoid rain is a weaker influence than the risk of injury during a gale, and both are much weaker than the risk of the death of your children.

The cases involve similar features, but to different degrees, which suggests that freedom of value-alignment admits of degrees.[95] Freedom seemed to be destroyed in the paradigm case of compulsion because the misalignment of motivated ends and valuable ends was forced upon the agent. In these new cases a restructuring is again forced upon the agent, but either the end the agent is left with motivation to perform is not as bad as in the paradigm case (it is perhaps annoying rather than abhorrent), or the agent's motivational bases for various ends is influenced to a lesser degree (they have good reason not to go outside, but no 'compelling' reason). So there are degrees to which freedom of value-alignment can be lost because there are degrees to which the agent is able to align their motivational bases for ends with what they value, and there are degrees to which the ends agents can contemplate compare to their conception of the good. Incidentally it seems fairly intuitive that compulsion comes in degrees, so this appears to be a sensible result to have arrived at.

What agents value and mesh-theories of agency

Now, those familiar with mesh theories of freedom will recognise this general approach. Mesh theories claim that free actions are ones that suitably mesh with features of the agent, and given this basic definition, freedom of value-alignment is a form of mesh theory. But why not adopt a well-developed mesh theory, such as Frankfurt's, that has been highly influential in the literature? We should note some of the differences then, and why the freedom of value-alignment may offer

[95] See O'Connor (2009) on degrees of freedom.

some advantage.[96] Frankfurt (1971) offers a hierarchical mesh theory according to which agents have free will when their action issues from a first-order desire that the agent wants. He models an agent's capacity for willing on a hierarchy of desires. First-order desires can produce actions, but need not do so. The agent also possesses second-order desires, the objects of which are first-order desires. If an agent possesses a second-order desire for a certain first-order desire, and if the agent acts on that same first-order desire, then the agent has acted with a will that they want, and has acted freely. For example if an agent has a first-order desire to eat a piece of cake, and another first-order desire to eat a piece of fruit (the two are in conflict), and a second-order desire to desire to eat a piece of cake, then if the agent's desire to eat a piece of cake produces the action of cake eating, then the agent will have acted freely. However, if the agent possessed no second-order desire to desire to eat a piece of cake, and instead possessed a second-order desire to desire a piece of fruit (perhaps they are on a diet), then the production of a cake eating action by a first-order desire to eat a piece of cake would be an instance of an unfree action. If an agent possessed no second-order desires relevant to either of these first-order desires (one to eat cake, the other to eat fruit), and the agent happened to act on one of the first-order desires, then the agent would not only be unfree, but would be a wanton, with no control over their actions, but passive to being moved, unreflectively, by their various first-order desires. According to Frankfurt wantons are not even persons.

One concern about this hierarchical theory is whether it captures the phenomenology of action and willing. In one way, it does powerfully capture something of our experience of what it is for have a will to act a certain way. I may desire a piece of cake, but wish I didn't desire cake so much, because I know it is bad for me while I am on a diet. This does seem to be an instance of possessing something like a second-order desire. I desire the end of my being constituted differently, so that I do not have such a penchant for cake. Note though, that it is a desire not to have a first order desire. Is this the same as having a desire for a different, the contrary, first-order desire? In this case it may involve the desire to possess some other desire, but the situation seems more complex than just possessing such a second-order desire, which indicates that, at least in this situation, having the will you want may relate to more than just the first order desire that produces action, and whether there is a second-order desire for this first-order desire. If I found myself desiring a healthy piece of fruit, I might be pleased with myself for having developed this good habit, or taste for fruit, and be glad, but is it the case that I possess a desire that I desire fruit? While I can be

[96] We will not do justice to developments of Frankfurt's ideas here, and my aim is not to assess Frankfurt in detail. However an alternative mesh theory is a useful foil for situating our discussion of freedom.

aware that I do not possess any conflicting desires, and have no reason to consider my desire to be bad, this is not the same as possessing a second-order desire whose object is my first-order desire. Rather than actively desiring this desire, I may instead be glad that I do not possess any conflicting desires instead. There may indeed be cases where I do possess a desire to have a certain desire, but it appears that we have reason to doubt whether what it is to have the desire that I want producing my actions is always captured by possession of a second-order desire.

I have attempted to sketch descriptions of agents who plausibly possess a mesh, but not a Frankfurtian second-order desire. I do not think that these descriptions are wrong, or reducible to Frankfurtian descriptions. Further, I do not think we would describe agents like this as wantons. We have not demonstrated that Frankfurt is wrong though, that would take a lot more work, and would be a considerable project in it's own right. But having done something to create doubt, let us compare with the alternative mesh-style analysis we have discussed. It does seem possible to account for agents' self-reflective attitudes to their own desires without necessitating a hierarchical structure. In terms of what agent's value, and their conception of the good, we can describe an agent who wishes that they desired fruit more than they do as an agent who desires the end of being constituted differently. The desire to eat fruit is not in itself valuable, but only instrumentally so, and it is the agent's desire to be healthy which motivates their adoption, promotion or identification with one desire rather than another. The agent's desire is not for a certain desire itself, but is that they be constituted in a way that aligns closer with their conceptions of the good and valuable. This is not a matter of possessing a hierarchy of desires regarding other desires, but is about the overall structure of the agent – what things do they value, and what are the relative values of the things they care about. Although agents possess desires that they would like to change, and lack desires that they wish they had, we do not need to make the possession of the relevant second-order desire necessary for freedom. These desires that are a bad fit, are a bad fit within the agent, they cohere badly with rest of the things that the agent values. An agent may, of course, be able to reflect on their beliefs and motivations, and express their judgement about their constitution in terms of a desire to change some facet of their psychology. But this need not be the case. Agents can act on desires that fit well, without possessing any desire for that desire. But agents like this are not wanton, and are still free in the sense that they possess freedom of action guidance.

But do such agents possess freedom of value-alignment? Well, this depends on how the agent came to be constituted as they are concerning their desires. If they have developed desires that have good and valuable ends, for the agent, as their objects, then they can be. This raises another problem for Frankfurt. Although the mesh of the hierarchy may give some insight into whether the action

the agent performs is one that expresses the agent's true self, the agent's true self could be one that was formed by manipulation. For example, remember our case of the hypnotist, who implants a desire to murder the spouse. If the hypnotist is clever, and implants a number of beliefs and desires about the benefits of murder, the harm of allowing the spouse to live, the desire to desire to murder, and so on, then a suitable mesh could be achieved. However an agent with this mesh would be the victim of a terrible, freedom destroying manipulation. Frankfurt does not have the resources within the hierarchy itself to explain why.[97] However we have successfully used freedom of value-alignment to conclude that an agent like this is not free.

The most pressing problem for Frankfurt is the regress problem. If agents can be wanton when they do not possess some desire that presides over their conflicting first-order desires, then why can this same problem of wantonness not occur at a higher order? For example, returning to our dieting agent who has conflicting first order desires about whether to eat cake or fruit, suppose that this agent also possesses conflicting second order desires – would the role of one of these second-order desires in the first-order desire's production of action not be something that is arbitrary? Even if there is not a conflict among second-order desires, how do we know that the agent has the second-order desire that they want? Adding third and fourth-order desires will not help, because we can always ask in virtue of what it is that the agent endorses the desires they possess. So if agents can be wanton with respect to their first-order desires because they lack a higher-order desire for that desire, then agents can be wanton with respect to all their desires. The hierarchy cannot be extended indefinitely, and so either an agent can possess a complex mesh, but ultimately remain passive and wanton with respect to their actions, or there is some other way of grounding a desire, or other action-relevant attitude, as one that the agent wants, and is active regarding. If an agent is ultimately passive regarding the structure of their will, or desires, then they do not have control over their means of action production. If the reason the agent has the will they have is, ultimately, arbitrary then the agent cannot be in control of their will. If the agent is not in control then something else is, even if it is blind luck, the random occurrence of a neural event, or any other candidate for what stands in the explanatory relation to the will the agent has. And, if something else is in control, then the agent is not free. The second-order desires introduced by Frankfurt were supposed to be the active exercise of control by an agent over their first-order desires by which an agent can 'identify' with one desire over another. Frankfurt later suggested (1987) that if an agent's higher-order desire is unopposed, then they do not have any need of further deliberation, and

[97] See Fischer and Ravizza (1998) for a discussion of this, though in the context of responsibility.

so no recourse to a higher desire is required. But, even an unopposed high-order desire can be one that the agent finds themself in possession of passively, and if this is so then it is not clear why the agent should identify themselves with any of the desires in their hierarchy. In this case, the desires lower down in the agent's will would be affected by something that has happened to the agent. Frankfurt amended the position (1994) to include a stipulation that the agent's identification must be active rather than passive, and when an agent has actively produced a desire then they are wholehearted towards that desire. However, we do not know in virtue of what an agent's production of a desire is active rather than passive, since the hierarchy was not able to secure it.

Watson (1975) saw the regress problem and tried to avoid it by making the mesh one that was not between different levels in the hierarchy of desires, but was between the agent's system of desiring, and their system of valuing.[98] However, as Velleman (1992) notes, the same problem will arise. If an agent can be subject to a desire operating upon him, then an agent can also be subject to a valuing playing a role in his psychology. There is no reason to think that the possession of a value is something that is more active than the possession of a desire. Bratman (2003)[99] offered an even more complex solution in which agent's plans and intentions are policies that guide his action production in a way that analyses the agent's identification in terms of long-term features of the agent's capacity for intentionality, and how their desires and valuings mesh with these policies. However, once again, the dogged question will arise: why is it that agents are not ultimately passive regarding the possession of a policy or intention? Velleman's own proposal (1992) is that instead of a motivational hierarchy, or mesh with another system, it is more of a matter of an agent's ability to identify with a desire by making it prevail by adding to it the motivational weight of the agent themselves. However the agent is identified as whatever it is that plays the role of resolving motivational conflict. This is not a higher-order part of the agent, nor another system parallel to the agent's motivational system. The picture is more like a web, and the influence of the web upon deliberations situated within it are the influences of the agent, and are those which Velleman would identify as active.[100]

---

[98] There is a similarity between Watson's position and the one we have been discussing, but there is not much to be gained by examining Watson's proposal in detail, the crux of the matter we are about to arrive at will be the same anyway (and having arrived at the proposed model in the previous section quite independently of consideration of Watson's theory, the similarities are incidental).

[99] Bratman's position is developed over a large body of work, and his 2003 article is a succinct articulation, but see also Bratman (1987, 1997 and 2004).

[100] There is some similarity to our previous discussion of freedom of value-alignment, which involved coherence with the agents values, and conception of the good. So,

However, no matter how we develop the agent's relevant features the question will still arise. Even if we give them a hierarchical structure of motivations, with parallel systems of valuing, all within a complex web that comprises the entire psychology of the agent, if the constitution of the agent is something that has happened to them, then they are not active, and have not identified actively with their desires, or ultimately their actions.

Now, note that this same problem attends the freedom of value-alignment. The proposal is a sketch, and there is much we could do to develop it, but we will not do that here. We have said enough to be able to identify when freedom is lost through external manipulation or influence – freedom of value-alignment does help us to identify freedom from this. But there is a further sense of freedom we need to identify. Complex though this proposal is, increasing complexity will not escape the problem of passivity. Our whistle-stop tour of some other mesh theories has taken us from the simplicity of Frankfurt's initial proposal (where the problem is readily identifiable) through to being able to see why it is not escaped by the freedom of value-alignment. Even if an agent is free from any external influence that would constrain their psychology in a way that prevented them from having a basis for self-determination that is aligned with what the agent conceives of as good, it is still an open question whether the 'self' doing the determining is something that was itself determined, or whether it was something that the agent possesses any control over. This brings us to the crux of the problem; if agents are going to be free then they need an active power of self-creation.[101]

Free action, and being up-to-the-agent

An agent may be instrumentally involved in their self-formation, and the on-going exercise of their capacities, but still fail to have ever had the capacity to make a difference to how that self is constituted – what they conceive of as good, what they value, and ultimately how they act. Now, in a sense, they have made a difference, for if the agent did not possess certain beliefs for example, then they would not have placed a certain value on an end, or been motivated in a certain way. However we can mean something else by 'making a difference', though articulating this has proven difficult for philosophers. One way to get at what we

---

unsurprisingly I am sympathetic to Velleman's insights. Ekstrom (2000) proposes a similar account in terms of preferences.

[101] I have borrowed this phrase from Wolf (1987), though she thought such a power was logically impossible. Kane (1998) has a similar concept, what he calls self-forming actions. However I have chosen not to use this term as it is defined partially in terms of voluntariness, and agents being causes of these action, and this associated baggage will be unhelpful to our analysis - so we can note the similarity, but I ask the reader to resist reading self-creation in Kane's terms.

mean is by analogy, the case of snooker balls is a classic example. During a game of snooker a red ball travels across the table with a particular velocity and spin, and goes into one of the pockets. These features of the ball make a difference to the ball going into the pocket, if the velocity had been different, for example, then the ball would have hit the cushion and bounced away from the pocket instead. But there is nothing that the red ball does that makes a difference to its possession of these properties; it is passive regarding the gaining of the velocity and spin it possesses. These properties are imparted to the red ball by the cue ball, which strikes it with a certain momentum, velocity and spin. So, what made a difference to the red ball's being so constituted as to go into the pocket was the cue ball. The influence of the cue ball on the red ball gave the red ball no opportunity to make a difference to whether it went into the pocket or not. A natural way to express this is that the red ball had no choice about whether it would go into the pocket or not. Once the cue ball had struck it, its fate was determined. So, the ultimate reason why the red ball went into the pocket rather than missed, is the cue ball[102], and not the red ball. And this is the sense of making a difference we are interested in. Notice it is not a matter of whether the red ball (or its properties) is the reason why it went in the pocket. In this sense the red ball did make a difference. For example, the red ball's position relative to the pocket is one of the reasons it went in. Instead the question asks for the ultimate reason that *X* occurred rather than *Y*, and is contrastive. It focuses our attention on the question of at which point the world went in a red-ball-going-in direction, rather than a red-ball-missing one. Once we identify this fork in the road, it asks why it happened, so we can identify where the buck stops for an explanation of the state of affairs' obtaining rather than not.

What would it take for the red ball to be able to make a difference? Well, it seems that the red ball must be able to do one of several things. It could avoid going into the pocket by changing the nature of its interaction with the cue ball. This could be by either changing the properties of the cue ball, or by changing the way that properties are acquired through a given interaction. Or, the ball could spontaneously swerve to the side, altering its properties of its own accord. The swerve could not be because of a gust of wind, or other influence, or it would be just like introducing another cue ball, it would have to be spontaneous in the sense that an occurrence of radioactive decay is spontaneous. An explanation can be offered that enables us to determine when radioactive decay is possible, but no explanation is possible that enables us to determine when, or that, an event of decay will occur.[103] Now to change the cue ball, the red ball would have to exercise

[102] It may be more natural to say 'how the cue ball hit it', but as will become clear, we do not lose anything by not using this locution.
[103] And this impossibility has nothing to do with epistemic limitations.

some power, the exercise of which could not be caused by another external factor, on pain of reintroducing the problem. To change the nature of the interaction the red ball would have to again exercise some power, or the nature of the interaction would have to be such that it admitted of some possibility of variation. This variation could not be something that was determined by other factors, otherwise we are merely introducing more 'cue balls', so it would have to be spontaneous. But introducing randomness as the reason why the interaction exhibits spontaneous variety, and hence a difference is made to whether the red ball goes into the pocket or not, will not help. Although a difference is made it is not made by the red ball.[104] So if the nature of the interaction is to be changed it must be by the exercise of a power of the red ball. These exercises of a power would plausibly mean the red ball spontaneously coming to have some new property, by which it either affects a change in the cue ball, which changes the nature of the interaction. So, however the red ball achieved the making of a difference about whether it went into the pocket or not, it would involve the ball in some self-change which was not determined by anything. So, we can see why the term self-creation is an appropriate one.

To extend the analogy to agents, we need only ask why the cue ball had the properties it had, which depends on how it was struck by the cue. This in turn depends on how the agent moved their body in using the cue to strike the ball. Did the agent make a difference to how the cue was used to strike the ball? Did they make a difference to how their body moved? Well, if the way the agent moved their body was the product of the influence of various 'cue balls' upon them, such as the lighting in the room, how much sleep they had the night before, how many hours practice they did this month, whether they had a beer before the game, their biochemistry, and so on, then what made a difference was these factors, and not the agent. Remember that we mean 'made a difference' in the special sense of being the ultimate reason why one thing happened rather than another. So an agent can be constituted so as to be able to perform actions, and can possess freedom of action guidance, and of value-alignment, but still not be able to make a difference to whether they act one way rather than another. If this is the case then the way that the agent acts is not up to the agent, and if it is not up to the agent, then it cannot be within their control. This is a different sense of control to that which we have discussed previously, and it would be a mistake to

[104] Such randomness happens to the ball, and passivity is not going to help us elucidate control. We will consider this issue more when we discuss the luck problem.

think that this power of self-creation was the basis for understanding the former sense of control as well, although they are obviously related to each other.[105]

Being up to the agent to make a difference about whether something occurs or not marks another sort of freedom then, which we will call freedom of choice. Unfortunately the term 'choice' is prone to come with a lot of baggage, which might make us think of particular types of conscious deliberation. However, we will adopt the term in the sense that the red ball was said to have no choice about whether it went into the pocket rather than missing, and questions about how the exercise of this freedom involves consciousness, deliberation, rationality, or any other capacities are a further question. We should note that this freedom of choice is incompatible with determinism. By determinism we mean that given the state of the world at some given time, there is only one possible way the world can be constituted at a later time, and that this later state depends, in all its details, upon the prior state of the world – because the states of the world evolve according to fixed natural laws. Facts about the past, the details of that past state, fix what the details of the present will be, and fix what the details of all future states will be. Consider our red ball again, whose velocity and spin were fixed by the influence of the cue ball. If we trace back the history of causal influences, then if at no point we find any spontaneous variation of any sort, then not only did the red ball not have a choice, but neither did the cue ball, and neither did the cue, and neither did the agent's muscles, and neither did the agent's motor cortex, their neural development between the ages of three and six months, their mother's diet during pregnancy, their great-grandfather's exposure to radiation working with Marie Curie, the Indo-European migration's effect on the population of western Europe, the formation of the continents through plate-tectonics, the formation of the solar planets following gravitational collapse of the pre-solar nebula, the evolution of the universe during the cosmic inflationary period, and so on. In short, there is no opportunity for there to have ever been any choice about anything.[106]

It so happens that this is not as much of a worry as is sometimes thought, since thanks to the influence of the Copenhagen interpretation of quantum physics, not many people really believe in this sort of determinism. The possibility of spontaneous variety in the outcome of quantum events means that the future states of the world are, to an extent, open.[107] However there is an extent to which

[105] There is considerable confusion in discussions of free will where it is assumed that there is only one sort of control relevant to agency, and only one sort of freedom relevant to whether agents actions (or wills) are free.

[106] We will avoid the thorny issue of the origin of this chain.

[107] I do not want to get drawn into a discussion about the interpretation of quantum physics, since this observation only mitigates the issue we are interested in, however, should the reader be concerned, it should be noted that genuine indeterminism also exists in classical mechanics as well, see Hoefer (2010).

the future can still fail to be open, as it pertains to its being open in regard to events over which no quantum events can make a difference. For example, if we put a gramme of radium in the red ball, there would be a chance of some radium atoms decaying between the cue ball striking it, and it going in the pocket. However with a half life of 1600 years, the change in mass of the ball due to the radium's decay would not be sufficient to make a difference to whether the ball went in the pocket or not. So even a form of nomological determinism qualified by the statistical interpretation of laws governing quantum behaviour will present a threat to the ability of agent to make a difference to things which we might think of as being up to them.[108] The things that we want agents to be able to make a difference to, are things like the value agents place on an end, the making of a practical judgement, and the actions they perform. There is no conclusive reason to think that these issues could be governed by the occurrence of a single quantum event, and much more reason to think that the physical processes involved in the production of these occurrences in agents relate to macro-scale physics. For example, it may be possible that some quantum indeterminacy in the behaviour of calcium ions at the synapses connecting effector neurons to a muscular system, means that there is some indeterminacy in the number of muscle fibers that contract. Although this may open the possibility of an agent's arm moving with a slightly different speed, it will not open the possibility of the agent's arm not rising at all. So although quantum indeterminacy opens some possibilities, whether the agent raises their arm (that is, what action they perform) is something that is still governed by determinism.[109] We cannot discuss the openness that quantum events may allow in the world any further here, but the point is that even this modified form of determinism is still a threat to freedom of choice.

Freedom, forking paths, and alternative possibilities

One way of articulating why freedom of choice is incompatible with determinism is to consider this freedom as the ability of agents to guide themselves through a garden of forking paths. Fischer describes the argument thus,

> We tend to think of the future as a branching, tree-like structure with many nodes at which there are various paths into the future. Perhaps there are points at which we have no alternatives to our actual course of action. And maybe

[108] There have been some attempts to ground the openness of agent's futures concerning their action in quantum events in the brain, however I am skeptical that this will be helpful. We noted that mere indeterminism is not helpful for securing a controlled freedom of choice, and given the stochastic nature of quantum events, control over their outcomes is ipso facto impossible.

[109] There is also a concern about whether such quantum events could constitute the exercise of control, but we will say more about this in due course.

> some individuals rarely or never have such alternatives. But we naturally think that there are many times when we (most of us) have various paths genuinely open to us. That is, we think of ourselves as frequently having alternative possibilities.
>
> When we take one path rather than another in a situation in which the other path is genuinely available to us, we say that we have a certain *control* over our behaviour. In this kind of circumstance, a person has the sort of control which involves alternative possibilities: he follows one path, and yet he *can* ("is able to," "has it in his power to") follow another path (1994: 3-4).

Now, there is some intuitive appeal to this analogy, because our experience as agents does seem to suggest that we do sometimes stand at points of decision where the future is genuinely open, and something we do makes a difference to how things turn out. However, this phenomenological argument has been undercut by the suggestion that epistemic concerns could also explain why we experience our sense of agency as we do. Dennett (1984) in particular has argued that if agents do not have epistemic access to the causes of their actions at the micro-level, because these causes leave no phenomenal footprint, then our experience would be precisely as if the future was open, when it may, in fact, not be at all.[110] So, by itself, this may be more of an intuition pump than a persuasive argument for incompatibilism. Nevertheless, it helpfully connects two important ideas: up-to-us-ness and alternative possibilities. Another way of talking about whether something is up to an agent, is whether they could have done otherwise. If something is not up to an agent then they possess no power over whether one possibility obtains rather than another. If there are no open alternative possibilities, then there is nothing that the agent can do, no power they can exercise, to leave matters up to them. In terms of the garden of forking paths, the picture would be of a path, tracing the agent's history, which never contains a fork. If anything is ever up to the agent, then this entails that at the point that something is up to the agent, there is some set of alternatives that the agent could possibly have made obtain through the exercise of their agency. But note that alternative possibilities only mean that there is something that can be up to the agent, and not that anything actually is – so alternative possibilities are necessary but not sufficient for up-to-us-ness (and hence freedom of choice). The existence of some alternative means that there is something for the agent to control, but whether, and how, that control is exercised is a further matter.

Alternative possibilities are often discussed in terms of whether an agent performs one, or other, of two actions, often with great moral implications

[110] See Goetz (2008), chapter 2, for a discussion of possible responses, and the limits of Dennett's claim.

attached to each.[111] But, the alternatives can also be a matter of whether an agent performs an action or not, so a relevant alternative can be refraining from performing a particular action, and not performing some other action instead.[112] This is particularly important since it is not uncommon for compatibilists to, perhaps rightly, charge incompatibilists with placing an unreasonable condition on free action.[113] If agents only have one reasonable course of action presented to them, it seems strange to assert that they are only free if they can also do something quite strange, and unreasonable, instead. Consider a case raised by Steward (2012), we are asked to consider Joe, who has to decide whether to move in with his girlfriend. Joe has many good reasons to move in with his girlfriend, emotional, practical, financial, etc. He lacks any reason to decide not to move in with her. It would then be strange to insist that an agent's freedom requires the possibility that they act in a way that contravenes their reasons. In the case of Joe it would be strange to insist that he was not free to choose to move in with his girlfriend unless he could also decide not to do so. Steward thinks that the compatibilists are correct about this, saying,

> an insistence upon such chances cannot help to secure or improve the agent's control over the course of events. I will argue that they [the compatibilists] are quite correct in thinking so and therefore, if the libertarian thinks that having the power to φ requires the existence of some objective chance that one will φ, she is mistaken, since when what puts one's φ-ing quite out of the question is only such things as one's own wants, principles, motivations, etc. (and where there are no further special worries about how these wants, principles and motivations have been arrived at) there should be no concern that an absence of possibility here amounts to a lack of freedom (2012:126).

Now, given our criteria for action, we might suspect that agents like Joe are rather unusual. If Joe really possesses no reason at all to refrain from deciding to move in with his girlfriend, he must be a position where he has no desires (or other motivationally relevant states) that do not aim, directly or indirectly, at his moving in with his girlfriend. If Joe did possess a desire that aimed at some other end, such as a desire for a cup of tea, then he could act on it, and take relevant features of the world as reasons to so act, such as that it would quench his thirst. Joe would not be deciding not to move in with his girlfriend, but would be refraining from making that decision, and acting in some other way instead. But

[111] In fact, there is a tendency for alternative possibilities to feature particularly prominently in approaches to the free will problem that focus on free will as a condition for moral responsibility, and approach which we are not adopting here.
[112] Although there is some debate about whether refraining is an action (or negative action), see McGrath (2005), Thomson (2003) and Vermazen (1985). See Muller (2012) for a good discussion of what could constitute the branching of a path.
[113] Although not a compatibilist, Steward (2009) notes that compatibilists are right about something when it comes to just how open the futures of libertarian agents should be.

perhaps we can place Joe in a position where refraining from either deciding to move in, or not, is not a possibility. Perhaps his girlfriend has cornered Joe, and delivered an ultimatum that he either tell her, or show her, that his is committed to moving in, or not. Now, going and making a cup of tea is tantamount to admitting his lack of commitment. But surely Joe could pause for thought, or to look deeply into his girlfriend's eyes, or to tell himself that he really must focus and stop thinking about tea. It looks quite difficult to place an agent in a situation where they do not have any alternative possibility open to them that is a viable action, one which engages their rational and agential capacities. This is no surprise, since we have already argued that agents' actions are not necessitated by their motivations, or their strongest motivations. Even if his girlfriend has set a time limit, and failure to respond by the set time will constitute a 'no, I will not move in', Joe could still fail to act, not from a conscious desire to wait and see what his girlfriend says (since, by stipulation he has no desire relevant to this end), or in order to achieve any other end, but just out of absent-mindedness.

However, we still might be interested in what Joe does regarding his response to his girlfriend in another sense. In many situations in which we are interested in freedom, the fine details about exactly how and when an action is performed, or not, are not of paramount importance. Joe will at some point soon tell his girlfriend that he is committed to moving in, or will fail to do so. But is Joe free to decide not to, and to wait until the time is up in the uncomfortable knowledge that he does not want to commit to moving in, even though he cannot come out and say it? When Joe fails to decide due to his absent-mindedness, this is different to an active decision to remain silent. It is not that his absent-mindedness is passive, and his failure is something that happens to Joe, robbing him of the ability to act, but forgetting to act is not something that agents seem to actively do – forgetting is not an action. Because of our interests in responsibility, and the significance of what Joe can do (what lies within his agential power given his constitution) and it's relevance to his relationship with his girlfriend, which is something valuable (despite its dysfunctional nature), we will often want to ask whether an agent like Joe was free to choose to move in, or *to choose not to*. We want to know whether actually deciding to move in is up to Joe, not just how, or when. Remember, up-to-us-ness is about the power to settle whether *X* rather than *Y* occurs. And if *X* is moving in, Joe may be able to settle *Y* (not doing so) by his inaction, but it would be more significant of a settling, both for Joe and his girlfriend, if Joe can settle matters *Y-ly* by his *action*. An active decision not to move in would involve some self-creation of Joe's psychological constitution, or character, that would have significant ramifications. Such an active settling of matters involves his intentionality, and other important features of his character, self-formation, and

agency. So Steward, and compatibilists, are still identifying a question that is important.

Alternative possibilities and the effects of freedom in the past

Given how Joe is constituted, he cannot choose not to move in with his girlfriend. So it seems correct to say that Joe is not free to decide to not move in with his girlfriend. If Joe makes a decision, how that decision goes is not something that is up to him, because there is no alternative to his deciding to move in with his girlfriend. So if the way that Joe can act admits of no alternatives, and alternative possibilities are a necessary condition for an agent having something up to them, and within their power of self-creation, then does Joe fail to possess freedom of choice? Well, it all depends on why Joe is so constituted as to be unable to decide otherwise than to move in with his girlfriend. To understand why, we can consider an analogy with ascription of culpability.[114] When it comes to ascribing culpability we are used to considering whether the way things are is because of an agent or not. It is not the only relevant question, responsibility is more complex than that, but we are sometimes led in our intuitions by a consideration of whether the obtaining of some state of affairs is something that was up to the agent or not. Once again, if it were up to the agent, then we would expect to find some alternative possibility. The usual case in these sorts of discussions concerns the drunk driver, who hits somebody with their car because they lack the requisite reaction speed to have any possibility of braking in time to avoid hitting them. Given the way things are with the inebriated agent, it is impossible for them to brake in time to avoid the accident, and there is nothing they can do to make braking in time a possibility.[115]

Of course, this would not be much a of an excuse for the driver, and were they to offer such an excuse, we would be likely to retort that the reason that they found themselves in the position they did was because of their choice to drink and drive, and that this was up to them. The driver is responsible for being in a position that is bad because of something they did that was up to them (we do not need to examine all the ways of thinking about this badness now – neglect, risk, foresight and so on). Notice though, that in this instance 'responsible' is not meant in a moral sense. What our intuition is picking up on is that the driver is agentially responsible for hitting someone with their car, because they are agentially responsible for putting themselves in the situation because of some prior

[114] Remembering that I do not wish to claim that freedom and responsibility amount to the same thing, or are even contiguous.
[115] Readers concerned about strange occurrences, like braking for other reasons and so on, can stipulate further conditions to secure the impossibility of any alternative.

action or choice. An agent is agentially responsible for some event when the determination of that event's occurring is something that is up to the agent and under their control – it is something they have freedom of choice regarding. Although the agent is not able to control the conditions that limit there being any alternative possibilities at the time of the accident, they were able to control the conditions leading to this limitation at a prior point.[116] So freedom of choice is something that can be derivative. The driver's failure to brake is up to the agent, because although he has no alternative possibilities at this point, his finding himself in such a position is itself something that is up to the agent. Consider another case where someone has cut the brake lines of the car. Now when the driver comes to brake he again has no alternative possibilities, but this time his finding himself in such a situation is not something that was up to him, it is something that has happened to him, without his exercise of freedom of choice, and is outside of his control.[117]

We can relate this to the analogy of the garden of forking paths. The driver finds himself on a path that is destined to include their hitting someone with their car. Because of the features of the path they find themselves on, their hitting someone is determined; it is an inescapable occurrence in the life of the agent. The agent does not possess the possibility of exercising their agency in a way that will create any alternative future regarding this detail. But we can trace the path back to an earlier junction with a path that provides the possibility of avoiding the accident. This could be because on the alternative path the accident is determined not to occur, perhaps because the agent gives their car keys to the barman, or the path may leave the accident's occurrence undetermined, and further junctions lie between this point and the time that the accident could occur. If the taking of one path rather than another from this junction is something that is up to the agent, then things that are determined by the choice of path from the junction are also up to the agent. At the risk of multiplying metaphors, it is like an agent who pulls on a piece of rope that runs across the room and is tied to a table, leading to the table falling over. The agent is agentially responsible for the table falling over. They may well say that they only pulled on the few inches of the rope they touched with their hand, but because the table falling over was determined by this occurrence, they are agentially responsible for the table falling over as well.[118] If their pull on the rope only pulled the table three feet across the room, perhaps away from the wall to being closer to a doorway, then things could be different. If some children

[116] Pink (1996:89) makes the point this way, and notes that this is considering alternative possibilities in a way that is autonomy-centered rather than time-centered.
[117] There are some similarities to Wolf's (1990) claim about 'bad' choices requiring alternative possibilities.
[118] If they could not see that the rope was attached to anything, then they might well fail to be morally responsible of course.

come running through the door, but are not determined to run into objects next to doorways knocking them over, then the agent is agentially responsible for the possibility of the table being knocked over, which needs to be distinguished from their determining that it is knocked over.[119]

Notice that I have not said that the future events, which are determined by the taking of a particular path from a junction, need to be foreseen, or even possibly foreseen by the agent. This might seem a little strange, making agents agentially responsible for things which may even be a long way in the future, or things which they could not have possibly foreseen. However I suspect that concern here is due to a lingering connection with moral responsibility. We do sometimes reflect on the past, and with hindsight note the significance of past choices given their impact on our future possibilities. We can be thankful that we chose a particular course, not because we were hoping to achieve the end we now value, but because it worked out that way. We can also regret not having chosen differently if we are prevented from pursuing some end because of past choices. But we do not praise or blame ourselves for these past choices, and friends might say that, "we couldn't have known", but they will not say, "that had nothing to do with you". Friends may say, "there was nothing you could have done", but in these contexts it seems this is shorthand for 'there was nothing you could have intentionally/knowingly done'. Agents may react with surprise to the effects of their actions, but it does not absolve them from playing a role in bringing them about.[120]

Although this seems to put an awful lot in the power of agent's past actions and choices, we should note that the garden of forking paths is rarely as simple as

[119] In this case, plausibly, the children may well be agentially responsible for knocking the table over, but the agent who pulls the rope may be morally responsible – highlighting the need for care in distinguishing the two types of responsibility.

[120] We might be concerned that we need to say more here about intentional action. Intentional actions are plausible candidates for our most significant actions, at least, in terms of expressing ourselves. But I do not want to get drawn here into a lengthy discussion of what kind of knowledge (and other conditions) are required for an agent's action to be intentional. Although linking significance to intentionality has some limited promise, I think that in terms of whether agents are free to perform actions that are free and significant, we have said enough with the freedom of action guidance, and freedom of value-alignment. The sorts of circumstances that could limit an agent's knowledge in a way that limit their ability to be intentional in performance of an action, whether personal interactions (like a manipulator) or natural, are already covered by these types of freedom. So if an agent was not free to act intentionally, we would already have identified them as not free to possess action guidance and/or value-alignment. The conditions for intentional action add a complication that will have to be left for another investigation, where space can be given to its subtleties (just consider the length and depth of Anscombe's excellent Intention (1957)). It is also plausible that there can be non-intentional actions that are significant, so there is some justification for taking this question to be tangential, or at least peripheral, to the main concern of this project.

these limited examples can describe.[121] We might be concerned that we have just 'pinned the blame'[122], so to speak, for all occurrences in the life of an agent, on their distant past choices. This would misrepresent the situation. Surely my past choices have influenced my life, but did not set my future in stone? Although there are choices agents can make at a junction that *determine* some features of the future path, there are a great many other influences on the evolution of the path that an agent takes. Indeterminism in the natural order (luck, randomness) and interaction with other agents (human and non-human) all affect the evolution of the path that traces the life of an agent. There can be some choices that do determine a great deal, perhaps of great significance to the agent. But often when we engage in this sort of reflection on our past choices matters are complex. Although my past choice may have made possible some future event on my path, that event finally obtaining probably depended on a great many other contingent junctions, at which my path may have diverged, and things may have gone differently. Sometimes, it would have taken some unlikely event, or a great many contingent events together, to have finally evolved my path away from some event occurring. So as I look at the possible paths further down the path, there may be fewer of them that allow an escape from the event's occurrence. So our 'pinning the blame' is sometimes probabilistic, as is our ability to act in a way that engages our plans for the future, which explains why we may feel an intuitive disconnect with the claim that freedom of choice can be derivative. However the garden analogy remains a useful one for talking about whether a particular degree of freedom, or freedom regarding some particular event, is something that is up to the agent.

So, returning to Joe, who finds himself unable to decide any other way than to move in with his girlfriend; Joe could not have done otherwise, and so how his decision can go is not something that is now up to him. But although he cannot now decide otherwise, he could (past tense) have used his freedom to be otherwise constituted now. If Joe had exercised his freedom of choice differently in the past, then he would have ended up with different reasons concerning his romantic attachments, financial security, practical living arrangements and so on. He may have actually found himself in different external circumstances, like having different job prospects, or may have had different ideas about what he wants and desires. If the reason that Joe now possesses no reason to not move in with his girlfriend is because of a past exercise of Joe's agency, one where he possessed freedom and there were alternative possibilities relevant to the possibility of his

[121] We should also note that some philosophers have suggested that agents may have relatively few opportunities to stand at a junction and have their future path be up to them, for example if matters are only up to agents in Buridan's Ass type situations, where a tie break is needed, then these situations are probably rare; others have disagreed.
[122] Not morally, but agentially.

presently being constituted with some reason to decide differently, then Joe's being only able to decide to move in, is something that is up to him. Note that it is not enough to just ask whether there were alternative possibilities in the past. It may well be that Joe possessed alternative possibilities in the past between being constituted with no present alternative by either being motivationally dominated by fear of rejection, or by emotional attachment to his girlfriend. While these two possibilities may be significant for Joe, they do not mean that whether he decides to move in with his girlfriend, rather than deciding not to, was ever something that was up to him.[123] This is where the garden of forking paths analogy is helpful: if something, deciding *M* rather than not-*M*, is up to the agent then they must now be at a junction where at least one open path is their deciding *M*, and at least one is their deciding not-*M*, or, there must be some past junction where at least one open path determines their later deciding *M*, and at least one determines their later deciding not-*M*.

So there is no need to make the rather counterintuitive proposal that in order to possess freedom of choice Joe must have the power to decide, against literally *all* reason, not to move in with his girlfriend. But for Joe's deciding to move in to be free, it must be the case that in the past Joe possessed some opportunity to avoid being left with no alternative possibility regarding his decision now – and whether this opportunity was taken or not, needs to be something that is settled by Joe. Joe may have had many opportunities to avoid being constituted as he presently is, but if these were all averted by the actions of his girlfriend (who happens to be an evil neurosurgeon) then how Joe presently decides is not something that is up to him. If Joe's girlfriend somehow engineers his psychology, not by implanting alien desires and beliefs, but by removing anything that gives Joe a reason not to move in with her, then Joe's decision is up to her, not up to Joe, so Joe is not free. Note that were she to interfere with Joe to remove some reasons not to move in with her, perhaps leaving only one, then Joe would still be free if he settles the matter and an alternative possibility exists for him. In fact, Joe's girlfriend could interfere with Joe at some point in the distant past, perhaps just after he speaks to her for the first time at a party, and could manipulate Joe so that he has one free choice later during that party that with either determine that he move in with her, or will not so determine. If Joe settles the choice between these two alternatives, then his decision much later to move in with her, the one that now lacks any alternative possibilities, is still free.

But surely there is something wrong with saying Joe would be free under such manipulation? Well, what if a choice Joe freely makes at the party, perhaps to kiss

[123] Note however, that although he may not be free regarding whether he decides to move in with his girlfriend, he would be free regarding why he decides to move in with his girlfriend.

the girl he has met, does determine that he will have no alternative possibilities years later when he has to decide whether to move in with his girlfriend. There is no manipulation – Joe's girlfriend is not an evil neurosurgeon, and has not planned to trap Joe, or manipulate him in any way. Would this trouble us in the way that a personal manipulator does? I do not think it does. We can even imagine Joe thinking about his decision to move in or not, and saying to himself, "from that kiss onwards, all those years ago, wheels were set in motion that caused me to fall hopelessly in love with this girl, so that I cannot help but move in with her" (apologies for Joe's rather cheesy internal monologue).[124] We do things all the time that curtail our future possibilities, and things happen to us all the time that have substantial impacts on our future possibilities. This does not mean that agents are thereby losing their freedom of choice. There remains a substantial amount that is still up-to-agents, because it is relatively rare for agents to perform actions that have such extreme impacts on them that they are left with no reasons to act otherwise whatsoever. Agents do have the power for self-creation, but they are also the product of their environment, and the myriad of factors that entails. They exercise their power for self-creation, for making themselves who they choose to be (the way they are constituted), in response to the challenges their interaction with their environment throws up – this power is not absolute, which is a very widely held admission. The fact that freedom of choice, so construed, allows for this complexity, and the possibility of who we are being affected by more than our own choices, counts in favour of this analysis.

However there is something to be said about Joe's girlfriend (the evil neuroscientist) when she interferes with Joe. There is a curtailment of Joe's freedom here, but not by taking away his power of choice. It brings us back to the beginning of this chapter, and to freedom of possibilities. We noted that things happen to agents that limit, or open up, future possibilities. We said that there might well be ethical guidelines relevant to thinking about how we impact on other agents. It might be bad when we impact others in a way that curtails the possibility of them pursuing projects that would be of worth to them. It might be good when we impact on others in a way that enables them opportunity to discover the good, and embrace it in their choices. So, Joe's girlfriend does curtail Joe's freedom of possibilities, and it is the ethical question this raises that provokes our reaction that this is unfair, and Joe is less free. We can see that this is an ethical judgement we are making about Joe's situation, rather than a metaphysical one about his powers of agency, by changing the moral status of the interaction and noting how we judge differently. Consider the case where Joe is pathologically afraid of being alone, and so cannot help but decide to move in with his girlfriend.

[124] There is a similarity between this way of arguing and Pereboom's four-case argument (2001). Dennett also discusses the pitfalls of personification of determiners (1984:61).

However, Joe's girlfriend won't let him decide until she has talked with him (she is now a psychoanalyst, not evil) and helped him to manage his fear and realise the open possibilities before him. He is no longer constrained in his choice, and is free to choose to move in 'for the right reasons'. I think we would now judge the manipulation to be something that, in the relevant sense, enhances Joe's freedom. This case opens up possibilities, so let us consider one more in which possibilities are curtailed. This time we will make Joe have reasons to move in with his girlfriend as well as reasons not to. His reasons to move in are good ones, he loves her, it makes financial sense, and so on. But his only reason for not moving is a pathological fear of abandonment due to an incident in his childhood. He knows, when he thinks about things clearly, that his girlfriend would never abandon him, and were she ever to leave, it would be due to a breakdown in their relationship, under which conditions he would not want to live with her anyway. So, he realises that the possibility of her leaving is not a reason not to move in with her. It might seem contrived to make Joe have this pathological fear as the only possible motivator for deciding not to move in, but Joe (as you may have realised by now) is quite a strange chap. Joe's girlfriend is once again a skilled psychoanalyst, and she influences Joe into losing his pathological fear. Joe is left with no alternative possibilities. But it is not clear that we should say that Joe's freedom was thereby destroyed.

The ethical questions raised by our impact on others do seem to concern an agent's conception of the good, as well as some objective judgement about what is good. When we interfere with agents in a way that allows them to align themselves more closely with the good, we are doing a good thing. When we divert people from the good, or do not allow them to form their own conception of the good based on their own exploration of matters (that is, we do not allow them to come to conceive of something as good by discovering the goodness of it for themselves), we do something bad. I think these are plausible principles, but do not mean them to be dogmatic. They are just examples of how we could think about the relevant issues. A useful example of these principles is that of the young initiate who is duped into joining a brainwashing cult.[125] The initiate makes a free (as in up-to-them) choice about joining the cult, but one that is made in significant ignorance. They are then subjected to manipulation, leaving them with values, and a conception of the good, that are at odds with ones they would have approved of prior to joining the cult, or ones that they would have discovered are actually bad, given the opportunity to freely investigate the matter (as in freedom of possibility). However, if we discovered that a loved one had been manipulated by a cult, and were sure that our judgements about their values and beliefs were well founded (perhaps we have received an infallible moral pronouncement from

[125] Kane (1998) considers a case like this.

God), then although our interaction with this person to manipulate them out of their brainwashed beliefs might involve a very similar type of manipulation in terms of altering their conception of the good, and their values, this would not only be good, but might be something that we have a duty to do. Although these ethical questions are interesting and practical, we will not dwell on them further here. However, it is interesting to note that God has a particularly important role to play in this regard when it comes to providence, and his interactions with human agents. Human agents may make choices that so constrain their future possibilities that concerning some things, it is now impossible for them to change. Such an exercise of freedom of choice to constrain future alternative possibilities is very significant for the agent concerned. But, there is nothing about such an exercise of choice that destroys the possibility of such an agent remaining free in the future. They can still possess freedom of choice, as well as the other types of freedom. Since this point will be of obvious importance to us, it is worth highlighting.

Frankfurt-style counterexamples, and the consequence argument

So, alternative possibilities are a necessary condition on an action's being free, although the alternative possibility need not be concurrent with the action, and may be prior to it. The central importance of alternative possibilities means that we cannot go much further without saying something about Frankfurt-style cases. We have already met these, and have considered whether agents like Jones, who is subjected to the counterfactual interventions of Black, are performing actions or not. We saw that there was reason to doubt that it was possible for an intervener to really secure Jones performing the action required in the counterfactual scenario, but if we put those concerns to one side for a moment, we can see that there are other reasons to doubt that Frankfurt-style cases present the problem they are sometimes taken to. The point of these cases is to present a counterexample to the claim that agents need to be able to do otherwise in order to be free. Frankfurt's original target was to show that agents do not need to be able to do otherwise in order to be responsible, but the issues are the same if we consider the problem in terms of freedom of choice requiring that agents have an alternative possibility available to them. In Frankfurt's original case Jones was considering whether to kill Smith, and Black was stood at hand ready to intervene to secure Jones' going ahead with killing Smith, in the event of Jones not doing so by himself. The idea was that Jones could not have done otherwise, and so there were no alternative possibilities, and yet in the event, Jones acted on his own, and was responsible for killing Smith (both morally and agentially, though Frankfurt focussed on the former). Frankfurt did not elaborate on the details of how Black

would secure Jones doing what he wanted. Obviously Black physically moving Jones' limbs would not work, as Jones would no longer be performing the action. Bribing or threatening Jones might increase the chance of Jones doing what Black want, but for the Frankfurt case to succeed Black's desired outcome needs to be completely secured. There have been a number of influential developments of Frankfurt's original case to fill in this gap,[126] and cases now often feature a neuroscientist who is somehow able to read every detail of Jones' neurophysiology, even down to the quantum level,[127] and has also installed devices that will kick in should Jones fail to act, or be activated by the scientist or one of his measuring devices. Vihvelin (2013:97 ff.) has usefully noted that the only methods open to Black involve either blocking Jones' deciding not to pull the trigger in some way, or pre-emptively initiating a trigger-pulling event should Jones be about to decide not to shoot Smith.[128] The blocking device is triggered by Jones starting to perform a non-shooting action, whereupon the device, sensing the beginnings of this action, prevents that action from being performed and replaces it with a shooting action by causing Jones to perform that action in some way. The pre-emptor device is triggered by a reliable indicator that Jones is about to act in a non-shooting way, whereupon it jumps in to supplant this eventuality by causing a shooting action instead.

The literature on Frankfurt cases is vast, and we cannot discuss all the issues it raises. But we should note some prominent reasons to think that they do not present a threat to our claim that freedom of choice requires alternative possibilities. First of all we might doubt the possibility of a pre-emption device. If Jones possesses a free choice, then he stands at a junction, and nothing about the past determines which way he will go. So the device cannot read anything about Jones prior to his acting that would tell it which way he was going to decide, because it is the nature of free choices to lack any prior determinant. The possibility of a pre-emptor device begs the question against an incompatibilist analysis of freedom, at least, against ones like ours, by assuming that an agent's actions are determined by their prior mental states. But we rejected this in the chapter on action. By the time Black can know what Jones is going to do, it is too late to prevent Jones exercising his agency freely between alternative possibilities.[129] Blocking devices do not have this problem though. They work by

[126] See particularly Mele (2006) and (1995).
[127] Despite the fact that this is ex hypothesi impossible because of the uncertainty principle.
[128] Vihvelin is not alone in noting this distinction, but has a clear discussion of it, and some interesting remarks about why Frankfurt's strategy is doomed to fail, even if on is not an incompatibilist of the sort we have developed. I will not discuss her argument here however, as we are really more interested in identifying the sense of freedom we are interested in, than arguing for it.
[129] We have allowed that prior free choices can fix future action though, so it might be thought there is some scope for these devices to read past intention formation for example.

changing the agent once it has been determined how Jones is going to act by his beginning his action. Consider the blocking case in terms of the garden of forking paths. The path splits at the point where Jones makes[130] his decision: one fork leads to his killing Smith, the other to his not doing so. But then Black's device intervenes, and on the fork that would have led to Jones not killing Smith there is another change in the path, one that now determines that Jones will pull the trigger. But by the time of the second junction it is too late. Jones has already exercised his freedom to do otherwise to make the forking of the path at the first junction something that is up to him. The intervention of Black's device can prevent Jones from succeeding in his action, but it cannot prevent Jones from having freedom of choice, because it can only intervene after the choice has been made. If it is objected that the device intervenes between the start of the decision and the completion of it, in order to prevent this conclusion, then more problems arise. If at this mid-decision point (we will suppose, for the sake of argument that there is one) the path has forked so that the device knows its intervention is needed, then effectively the decision is made, even if Jones is not consciously aware yet, or has not moved his body yet. If the path has not yet forked, then it is true that the decision is not yet made, but it is also true that the device must now be functioning as a pre-emptor, and the former objections will apply. So a blocking device cannot rob Jones of a relevant alternative possibility.

There is a lot more to be said about how to respond to these cases (and learned from them). But this identifies strategies for responding, which is enough in the context of this project. However these responses do depend on the conception of freedom of choice as a particular kind of up-to-us-ness, and it would be good to provide some argument in its favour. We will be brief, but that is justifiable, because this plank in our theory of free action is one that we have to assume in the context of our theological puzzle. Agents need to possess at least some genuine aseity, which means that at some point there is a fork in the path that is entirely up to them, one that cannot be determined by prior states along the path. If it were not, then there would be no aseity, which is essential to the response to the problem of evil, whose coherence we are interested in. The consequence argument is the classic formulation of this concern that freedom is incompatible with determinism. The literature on this argument is vast, but thankfully we only need to consider it enough to clarify further the sense of making a difference, or up-to-us-ness that we are interested in. The locus classicus

---

However, in this case there is an alternative possibility for Jones still, but it is one that was exercised in the past, and nothing about Black's devices can now affect it. These types of scenarios are not an apt target for Frankfurt-style arguments.

[130] Or begins to make – though whether making a decision is an extended process possess parts is itself suspect, however we will let that pass for now.

is van Inwagen's *An Essay on Free Will* (1983),[131] in which he gives the basic form of the argument,

> If determinism is true, then our acts are the consequences of the laws of nature and events in the remote past. But it is not up to us what went on before we were born, and neither is it up to us what the laws of nature are. Therefore the consequences of these things (including our present acts) are not up to us (1983: 56).

The disagreement between compatibilists and incompatibilists centres on whether to understand the power to do otherwise in a conditional way or not. The compatibilists favour the conditional analysis, which allows the power to do otherwise expressed in 'could have done otherwise' to be analysed as 'would … if …'. For example incompatibilists might claim that Joe could have done otherwise than decide to move in with his girlfriend (and that this is important to assessing whether he was free). The compatibilists can respond that they agree, and what makes Joe free is that had he been constituted differently, with different desires and motives, then he would have decided differently than he did, which all goes to show that how Joe decides is up to him. But note that the compatibilist means that it is up to facts about Joe, whereas the incompatibilist means that it is up to how Joe exercises his power of agency. The compatibilists mean that had the past been different, Joe's decision could have been different. Perhaps the past could have been different, either because the laws are not fixed, or because there is natural indeterminism in the evolution of the world. But neither of these things are under the control of Joe, and especially not when they are located in the past. This conditional analysis captures the idea of being free from external constraint or impediment. But we have captured this with freedom of action guidance and value-alignment – in freedom of choice we have identified a different sense in which agents are free. Kane (1998:48) notes that it is typical for the discussion to stall when incompatibilists charge the compatibilists of begging the question when they import a conditional analysis of the power to do otherwise into the consequence argument. Compatibilists charge incompatibilists with begging the question by ruling out such an analysis from the start.

Kane also discusses several objections to the conditional analysis.[132] Austin (1956) suggests the example of standing over a three-foot putt. It is possible to make the putt, and it is possible to do otherwise – to miss it. However there is nothing to suggest that I would make the putt if I tried. Kane notes that therefore 'would … if …' statements are not a necessary condition for the truth of 'could…'

[131] Note also the important exchange between Lewis (1981), and van Inwagen (2000), (2004) & (2008).
[132] Since Kane provides such a succinct summary of the debate, I have followed him closely concerning the conditional analysis.

statements (54). He also notes that determinate causation is not required for agents to count as the cause of an event. The golf example is a case of probabilistic causation, if I try really hard, I may be more likely to make the putt. Kane mentions that Anscombe (1971) states that one can be said to 'cause' or 'produce' and outcome without that outcome being determined or inevitable.

Lehrer (1964) suggests another example in which the agent is presented with a tray of red sweets. Nothing prevents the agent from taking one of the sweets and eating it if they choose to do so. However suppose that the agent has a pathological aversion to things that are the colour of blood. In such a situation the agent cannot eat the sweets. However it is true that they would if they tried. Therefore the 'would … if …' conditional analysis does not provide sufficient conditions for an analysis of 'could do otherwise'. Kane discusses a supplement to the conditional analysis of 'would … if you had willed to choose …'. However we would then need to add the condition that the agent could have so chosen. Therefore the supplement is inadequate because it introduces another instance of 'could', or 'can'. Kane notes Davidson's response (1973) that the supplement of 'would … if you had wanted (desired) otherwise' avoids the regress to a 'can' because wants are not actions that the agent performs. However 'if you could not have done anything to control the aversion or overcome it, then you could not have eaten the sweets' (58). So the power to do otherwise and eat the sweets depends on the power to control a motivation. This is not to turn wanting otherwise into an action, but is to ask whether there was some action that could have been performed that would have made that motivation less effective in producing an intentional choice. This is again a matter of looking into the past of an agent and asking whether they exercise the power of self-creation to influence who they are. The important issue is how we acquired our psychological constitution. Which brings us back to our discussion of self-creation and up-to-us-ness. It may well be that compatibilists would disagree with this analysis, and I suspect that at this point the dialectic would stall, as it appears to have done in the literature, with both sides' arguments running against differing intuitions about free will.[133] I will not attempt to resolve this stalemate here, and luckily we do not need to, since we have theologically motivated the assumption of our intuition, and are investigating it's coherence. What we need to consider next is how this indeterministic power of self-creation could be coherently realised in an agent, for

[133] It is worth mentioning that more nuanced approaches to a compatibilist rendering of being able to do otherwise are possible, such as the new dispositionalists (see Vihvelin (2004) and (2013), and Fara (2008). However there is good reason to think that the dispositional analysis only explains how agents are in control of how they do actually act, and possible the possibilities for action (see McKenna (2009) and Fischer (2008)). But this has to do with freedom of possibilities and freedom of action guidance.

it will be how agency is embodied in an agent, and the mechanisms by which it is exercised, that are the possible focus of divine influence.

4. Agents, Causes and Luck

Having developed a sense of what agents would need in order to be free, we now need to consider how it is that agents could possess this freedom, and crucially how it can be exercised in a controlled way. Since, for our purposes, we are assuming incompatibilism (what we introduced as aseity), we do not need to argue that our analysis of freedom of choice is necessarily correct, or is true, but only that it can be coherently conceived. If it can be coherently conceived, that is, if a coherent analysis is possible that responds to prominent objections against this sort of incompatibilism, then we can trust our conditions of freedom and apply them to our theological puzzle.[134] Approaches to answering these questions by incompatibilists fall into three broad categories: event-causal, agent-causal, and non-causal. Event-causal accounts claim that free actions count as being controlled by an agent in virtue of the causal structure of the production of that action. Agent-causal accounts are similar, but rather than examining the causal structure of an action in terms of it's production by events, they adopt a substance causal approach, and claim that the agent must stand in relation to the action as substance-cause. Non-causal accounts deny that causation is the appropriate tool for analysing when an agent controls an action. There are a great variety of theories that have been developed within each of these classifications, especially concerning the processes involved in the production of actions, and the place of indeterminism in them, but we will not attempt to survey them. However, some recurring issues do arise. The most prominent issue for incompatibilists is the luck objection.[135] We will sketch this, and look at how it affects a couple of prominent theories. We can then consider a strategy for meeting this challenge.

The luck objection introduced

The luck objection is really a family of objections, but the common component is that luck, or randomness, or chance, is a control-diminishing factor in the production of action.[136] A helpful way to sketch the luck objection is to consider where we could locate indeterminism in action production.[137] By indeterminism we mean to pick out the possibility of doing otherwise, the point at

[134] Of course, our discussion, while falling short of establishing the truth of a certain analysis of freedom of choice, is one that I think is correct, and should go a substantial way towards motivating this analysis. Unfortunately, a thorough discussion of all of the pertinent issues is beyond the scope of this project, so we will only hit the highlights.

[135] It should be noted that the luck objection is also a problem for some compatibilists, see Beebee and Mele (2002).

[136] See the seminal Nagel & Williams (1976), and for overviews of the luck problem see Levy (2011) and Mele (2006).

[137] This approach is suggested by Ekstrom (2000).

which there is an alternative possibility (in a non-counterfactual sense). Let us assume a process of action production that is causal in nature, to see how luck may present a problem for it. Suppose that agents come to possess certain reasons, motivations and attitudes, and that there are events of their so doing, these then serve as inputs into a process of rational deliberation, which issues in a practical judgement about possible actions an agent could perform, the agent forms an intention, on the basis of their deliberation, which leaves them with a persisting motivation to perform a particular action which they have now chosen, then, when the conditions pertinent to that intention are satisfied, the agent performs the action. So, what if an agent is at the end of this process with a persisting, well-founded intention to do *A*; does introducing the possibility that they do *B* instead help establish their freedom? It seems that this is tantamount to insisting that agents are free when they can orient their agency towards one end, *A*, and then without further deliberation, do *B* instead. In such circumstances, we might worry that their doing *B* is an irrational act, since it does not represent the agents pursuit of an end which they have 'chosen'. Quite where we should identify choice is controversial, and perhaps a case can be made that agents form intentions, but that their choice is exercised when they finally act. But, for most, it seems that some choice has been exercised in taking up an end in an intention to act a certain way. The agent has done all the hard work of determining which end aligns best with their beliefs, desires, values and motivations, has taken into account the chances of success of various actions for achieving the end they want, and has committed themself to a course of action, to doing *A*. Their subsequent failure to perform *A*, and to perform *B* instead, would seem to be something that would come as rather a shock to the agent, defeating their exercise of agency-related powers up to that point.[138]

So, we can see why the language of 'luck' and 'chance' is appropriate here. If the agent does actually succeed in performing *A* rather than *B*, they could count themselves to be lucky to have had one of their intentions successfully issue in an action they wanted, or had chosen, to perform. If they fail to do *A*, then they might count themselves unlucky, and actually a victim of chance. After all, they had done everything possible to secure doing *A* (which perhaps was morally, or prudentially better), only to have the introduction of the indeterministic possibility of their doing *B* thwart their powers of agency leading up to that point. This seems to be a control-diminishing sort of indeterminism, because the agent has exercised their control in forming their intention to *A*, but then has no further control over whether they actually perform *A*. Although the agent has the power to introduce the possibility of their *A-ing*, their actually doing so, remains a further question. Wouldn't it be better if agents' intentions invariably guided the actions

[138] This sort of luck is discussed with the nice example of promise making in Mele (2004).

they perform – if their intentions determined the actions they perform? This is a plausible suggestion. In fact, the intervention of indeterminism seems akin to the intervention of a manipulator in depriving the agent of freedom of action guidance. Of course, randomness should not to be personified as another agent, acting upon the unfolding process. We rightly identify an event as random, when it lacks anything that determines its outcome. But this is the point here. Should the agent do *B*, and have their intention thwarted, there is nothing we can identify as the cause of the agent *B-ing* rather than *A-ing*. So, the ultimate cause of the agent's action is – well, there is no cause. It just happened that they did, after all *A*, or, it just happened that after all that preparation to *A*, they *B-ed*. After our discussion of the importance of up-to-us-ness, it is easy for us to see why freedom of choice prohibits this sort of indeterminism in the production of free actions. This sort of indeterminism does not diminish the freedom of agents; it destroys it. The fact that a question remains about the doing of *A* or *B*, means that the question of up-to-us-ness has not been settled, and in the absence of some mechanism enabling the agent to cause *A* or *B*, this question cannot be up to them.

So, what about locating the indeterminism in between the deliberation and the intention formation? We might be suspicious about whether there is any space between these two for indeterminism to occupy, but supposing that there is, it looks to be problematic. At least now the agent's intentions are not frustrated by chance, and so the agent has control over their actions because there is no gap between their intentions and actions. But the question of control now arises over the production of the agent's intentions. If they deliberated about *A* and *B*, and formed a practical judgement that *A* was preferable, should they then come to hold an intention to *A*, they would be lucky to do so. Contrariwise, should they come to hold an intention to *B*, they would the unlucky victim of randomness. It appears that for the agent to be in control there has to be a close connection between their deliberation and their intentions, just as there should be between their intentions and actions. So, if we are going to find a place for indeterminism, we will have to look at the process of deliberation itself. The agent has various inputs to this process (desires, motivations, etc.) and concludes the process with a practical judgement. Given that our folk psychology and phenomenology of free actions often conceives of the process of deliberation as one that could go any number of ways, and where uncertainty is resolved, this may be a more plausible place to locate the indeterminism. But if we break the deliberative process down into a series of causally connected events and states, then think about what sort of connection we would like to obtain between them for the process to be one that is controlled, the same concern about luck re-emerges. For instance, if D_{AB} is the event of the weighing of the ends *A* and *B*, and D_A is the state of realising that *A* outweighs *B*, then we want there to be a close connection between D_{AB} and D_A. If

there was not a close connection then we might worry that despite the inputs to the deliberative process, an agent could produce any old practical judgement. It would be analogous to possessing a calculator that produced a random number, no matter what the sum entered was.

Perhaps we can place some limits on the possible practical judgements agents could produce by their deliberating, so that agents deliberating about what to have for lunch do not decide to bite the person next to them (though if there is no way to stop the slide into randomness, this could be a concern). But in matters of moral and prudential importance we often find agents faced with significant choices between two or more alternatives, such as which job to take, which person to save, or whether to move in with their girlfriend or not. If there is indeterminism in the process of deliberation then we might worry that agents' strongest reasons do not win out, which would be significantly at odds with our ordinary conception of deliberation as agents trying to figure out what they ought to do based on what they have strongest reason to do.[139] The last place to locate the indeterminism would be in the acquisition of desires, beliefs and other psychological states that can stand as inputs for agency. Beliefs, in particular, are normally taken to be things that agents acquire passively. If that were correct, then indeterminism about belief acquisition would be troubling. It would give us reason to doubt that those things that produce beliefs in us are well equipped to connect the content of our beliefs with those things in the world that our beliefs are about. Even if these inputs are not passive, and there are things, perhaps like desires, that we actively take to be desirable to a certain extent, then we need there to be a certain connection between our perceiving the features of something, our reflection of the value of those features (and perhaps our conception of the good). Luck would undermine our ability to form desires for things that we find satisfying, to believe in things we have reason to think true, and to take to be reasons things that connect us to the world, including, self-reflexively, ourselves.[140] Randomness limits our ability to possess this connection. It is analogous to Kant's veil that limits our access to the noumenal world of things as they are. Our practical and moral deliberations could be based on phenomenal representations of what is true, desirable, rational, etc., without the ability to secure that these representations resemble the noumenal at all. So we see that luck presents a potent

[139] I am sure that readers will find my causal characterisation of deliberation rather at odds with earlier comments about whether agents have to act on their strongest reasons, or whether they are caused to act by their strongest reasons. We will have more to say about this in due course, but for now I am only interested in sketching matters in a way that highlights ways in which the luck objection can manifest itself, and this happens to be more easily seen on a causal theory of action.

[140] We are ignoring for simplicities sake many substantive questions about the status of things like desires and reasons, especially internalism/externalism. But, again, we are only interested in seeing how luck could present a problem.

challenge. The insistence on some indeterminism, some place where matters are not determined by what goes before but can go otherwise, creates the possibility of things being random, or chancy. And, so the objection goes, this randomness undercuts the agent's ability to exercise control.

Event-casual responses: Kane and Ekstrom

Let us consider one of the most influential event-causal attempts to respond to this problem. We will look at Kane's account of free will, which is representative of event-causal strategies. Kane's account targets an agent's possession of free will, rather than free action, and is framed in terms of the freedom required for moral responsibility. So we need to say a little about the difference, but we will see that the essence of incompatibilist freedom as the up-to-us-ness required for freedom of choice, remains the same. Kane proposes a definition of free will; 'the power of agents to be the ultimate creators (or originators) and sustainers of their own ends or purposes' (1998:4). Kane thinks that while incompatibilists can accept compatibilism between intentions and action, they must insist on indeterminacy between reasons and choices. With these points in mind Kane suggest the following definition of willing and voluntarily willing (1998:30):

> (W) An agent *wills* to do something at time t just in case the agent has reasons or motives at t for doing it that the agent wants to act on more than he or she wants to act on any other reasons (for doing otherwise).
>
> (V) An agent acts *voluntarily* (or *willingly*) at t just in case, at t, the agent does what he r she wills to do (in the sense of W), for the reasons he or she will to do it, and the agent's doing it and willing to do it are not the result of coercion or compulsion.

These two definitions exclude coercion, and preserve intentionality. The distinction between willing and doing so voluntarily is important when considering problems like Frankfurt-style counterexamples where agents will something but only at the behest of another agent.

Kane suggest that freedom required for moral responsibility has two conditions: alternative possibilities and ultimate responsibility. He gives the following definitions for these two conditions (1998:33-35):

> (AP) The agent has *alternative possibilities* (or can do otherwise) with respect to A at t in the sense that, at t, the agent can (has the *power* or *ability to*) do A and can (has the *power* or *ability to*) do otherwise.
>
> (UR) An agent is *ultimately responsible* for some (event or state) E's occurring only if (R) the agent is personally responsible for E's occurring in a sense which

> entails that something the agent voluntarily (or willingly) did or omitted, and for which the agent could have voluntarily done otherwise, either was, or causally contributed to, E's occurrence and made a difference to whether or not E occurred; and (U) for every X and Y (where X and Y represent occurrences of events and/or states) if the agent is personally responsible for X, and if Y is an *arche* (or sufficient ground or cause or explanation) for X, then the agent must also be personally responsible for Y.

Alternative possibilities (AP) is a common condition in the free will literature, but Kane thinks that incompatibilists stake too much on this alone. Ultimate responsibility comes in two parts, and the ultimacy condition identifies the agent as the ultimate and originating cause. Kane mentions Aristotle's concern that agents must be originators, or must provide the explanation for what they originate.[141] He is thinking of something like character formation. If we are responsible for action flowing from a virtuous character, then we must also be responsible for the character we possess. In other words we cannot be only an instrumental cause. For an agent to be ultimately responsible the buck must stop with them. The responsibility condition identifies the agent as a necessary causal element for their responsible action. The agent's action (or inaction) does not provide sufficient conditions, as there may be other contributing conditions outside their control, for example other agents, or environmental factors. But in a case of responsible action, when we ask why the action occurred, our answer will have to include 'because of what the agent voluntarily did' in the sense of (V), and hence did not have to do. So when action is responsible the buck stops with the agent, and specifically with the agent's choice. This is very similar to our condition of up-to-us-ness in freedom of choice.

Kane responds to Dennett's (1984) challenge to (AP) in considering the case of Luther who uttered, "here I stand, I can do no other" and yet despite his self-exclaimed lack of alternate possibilities is counted as responsible for his actions. Kane argues that in order understand a case like Luther's we need to investigate how the person came to be the way they are. Was Luther responsible for being the kind of person that uttered as he did? If there was nothing that Luther could have done that would have made a difference to who he was then he would not have been responsible. But Kane notes that this is to invoke (UR), so in order to preserve the requirement for (AP) it has to be understood alongside (UR). Recall that (UR) has two parts, (R) contains the requirement for APs and that (U) contains the requirement that the agent be responsible for the *arche* or sufficient ground of action. Kane claims that it is (U) that captures the incompatibilist intuitions that cause them to object to compatibilist accounts (like Frankfurt's). Kane states that (U) makes it clear that 'ultimately responsible agents must not

[141] Compare with Chisholm (1964) and our discussion in chapter one.

only be the sources of their actions, but also of their will to perform the actions' (1996:73), (U) precludes the possibility of determinism if free actions exist. While heredity and environment limit possibilities, we want to know whether after all this is taken into account is there anything left over that is up to us? Kane proposes the term 'self-forming actions' (SFAs) for the kinds of acts an agent needs to originate if there is to be responsible action. He defines SFA's thus:

> SFAs … are the undetermined, regress-stopping voluntary actions (or refrainings) in the life histories of agents that are required if U is to be satisfied, and for which the agent is personally responsible in the sense of R. The agents must therefore be responsible for them directly and not by virtue of being responsible for other, earlier actions (as would be required if they were not regress-stopping). This means that, for SFAs, the "something the agents could have voluntarily done (or omitted) that would have made a difference in whether or not they occurred" is simply *doing otherwise*, rather than doing something *else* that would have causally contributed to their not occurring (1998:75).

So SFAs satisfy (AP) in an incompatibilist sense that requires indeterminism. Coming full circle to Dennett's challenge of Luther's, "Here I stand" brings Kane to describing the important thing about Luther being that his will was a will of his own making. When he acted as he did, he acted of his own free will. This ties in (UR) with this common phrase used when talking about free will and responsibility. Kane states:

> While SFAs are required by UR for there to be ultimate responsibility at all, SFAs *are not the only actions for which we are ultimately responsible*, according to UR. Luther's "Here I stand" would have been an affirmation for which he was ultimately responsible, even if it was determined and even if he could not have done otherwise, so long as it was a willed action (issuing from his character and motives) and he was responsible (as U requires) by earlier undetermined SFAs for the character and motives from which the affirmation issued. In other words incompatibilists do not have to hold that all actions for which agents are ultimately responsible must be undetermined; ultimately responsible actions form a wider class of actions than undetermined SFAs (1998:77).

Kane proposes that different actions can arise from the same past, without occurring merely by chance, by arguing that SFAs involve agents in parallel processing. Consider one of his examples[142]: a woman is on the way to a meeting that is important to her career, but while walking notices a mugging taking place in a nearby alley. She then experiences an 'inner struggle' where her moral conscience tells her to stop and intervene to save the victim, but her drive for

[142] Kane has used this in several places, I have adapted it from Kane (2011).

career ambition tells her that she cannot miss this meeting, and must keep walking. 'She has to make an effort of will to overcome the temptation to go on to her meeting. If she overcomes this temptation, it will be the result of her effort, but if she fails, it will be because she did not *allow* her effort to succeed' (2011:387). The woman possesses conflicting reasons, or motivations, and the presence of this conflict presents an obstacle to either of those reasons being effective in production of an action. Because the woman has a reason to stop and help, she tries to do the moral thing, and makes an effort of will to resist the temptation to pass on by. However because she has a reason to go to her meeting she also makes an effort of will to resist the temptation to delay going to her meeting by stopping and helping. Both of these efforts of will are of indeterminate strength – and it is here that Kane isolates the place of indeterminism. One of these efforts of will 'wins', and our agent acts one way or the other. Given the indeterminate strength of these two efforts, and their direct causal connection to the actions they produce, there is genuine indeterminism about which action is performed. Whichever action is performed, it will be true that it is an action that the agent wanted to perform, tried to perform, and had reason to perform. So, Kane claims, the agent is actively involved in the production of the action. Their production of the action is grounded in their character, and the motivations it can produce.

But, Kane has not escaped the luck objection. Let us focus on whether the agent has control over whether she stops to help, or goes on to her meeting. The agent possesses two conflicting motivations, and the agent wills something if, when they act, they do so in accordance with their strongest motivation. It appears that Kane wants us to imagine the agent's action producing faculties as being in stalemate when these conflicting motivations are at parity. It might be difficult to say that they need to be exactly equal, but we can rectify that by stipulating that motivations need to achieve a threshold majority to be ones we can identify as the agent's will. The efforts of will are not well defined, but since they act on the agent's motivations they seem to point to a hierarchy, or perhaps web, of motivations. There are lots of issues for Kane here, about whether these efforts of will are actions (in which case we might worry about a regress), about problems of regress in hierarchical models of the will, and about whether we can actively change our motivations by dint of effort. But even if these issues can be resolved, the effectiveness of these efforts of will is not something that is up-to-the-agent. Kane explains the inherent indeterminism of the strength of these efforts as being due to quantum indeterminacy that is amplified by the chaotic nature of complex neurophysiology. This fixes the final effective strength of the effort of will as being something that cannot be determined by any prior condition, and is genuinely spontaneous (within a certain range).

If an agent only had one effort of will, with determinate strength, then we could say that the exercise of this effort of will on the agent's motivation, in order to direct action production, constitutes an active engagement of the agent in controlling their actions. But our next question would be about whether the agent possesses control over their possession of a certain effort of will's being a part of their psychology. This is to apply consequence argument style reasoning to the agent's acquisition of, or possession of, this effort of will. Note that for Kane, the agent possesses an effort of will to resist a temptation, or do a certain act, because of their character. So with one effort of will, our focus would be on the regress to the agent's early character formation, and whether they exercised control.[143] We can apply the same consequence argument style questioning to the change of state of the agent that takes them from possessing two indeterminate strengths of effort of will, to the state where one effort is determinately stronger than the other. On pain of regress this change cannot be one that is caused by some prior effort, or change in the agent. Kane has located the root of this change in quantum indeterminacy to avoid just such a regress, or determination by some prior event or state. So, to the question of what the resolution of these two efforts of will is up to, the only answer that can be given is, 'chance'. Since control resides, for Kane, in his story about how agents cause their actions, this means that which effort of will is successful is something than cannot be under the agent's control.

However Kane has done something to explain an agent's control over their actions. The action produced when our agent finally walks on to her meeting, or stops to help, is one that fits with her character, is one that she willed voluntarily, and possessed an alternative possibility concerning. However it was not something that was up-to-her. She did not possess freedom of choice about whether she stopped to help, or went to her meeting. The complex story about her motivations, and efforts of will, does secure something, but it only secures freedom of action guidance.[144] But we have found that this sort of freedom is quite compatible with determinism. Therefore what Kane has done is to explicate a sense of freedom that is compatibilist in nature, because it relies on a compatibilist friendly source-model of agential control, but has developed his theory in a way that is sensitive to incompatibilist intuitions about the need for alternative possibilities.[145] However this does nothing to secure freedom of choice, and the introduction of indeterminism to the essentially compatibilist friendly story, has only created a problem of luck. We can appreciate the concern of compatibilists,

[143] I am doubtful about Kane's ability to contend with this problem, but we have no space to discuss it here.
[144] Kane's conditions of (V) and (W) also secure freedom of value alignment, the freedom from manipulation.
[145] Clarke (2003) has also noted that there is something 'compatibilist' about Kane's account.

who see an admission that control has to do with causation, and then find the possibility of a complete causal explanation of how the agent acts being undermined by an event which does not have a contrastive causal explanation. Kane does talk about the agent being an *arche* (a beginning point) for their action. But the *arche* he has in mind is the agent being the primary sufficient cause of the conditions from which action was possible, which secures the agent as the source of their action. However the *arche* we are interested in is that of a junction in the garden of forking paths, the *arche* at which the production of action went one way rather than another. There seems to be a deeply held intuition among many incompatibilists (especially those of a more agent-casual persuasion) that this *arche* should be the agent, or be caused by the agent[146]. Kane enables us to cite the agent as the reason why one of either *A* or *B* occurs. But the agent cannot be the reason why *A* rather than *B* occurs, or vice versa. On Kane's account the answer to this second question must be 'luck'.

Ekstrom (2000) develops an account that attempts to deal with worries about luck in an agent's process of deliberation, by placing the indeterminism at an earlier stage of the process. Indeterminism within the process of deliberation (or any other psychological process contributing to decision and intention) would mean that the agent is not really exercising control over the output of the evaluations their psyche makes. Instead the indeterminism must be located prior to intention formation. The key idea is that not all causation involves necessitation. In indeterminate causation the cause merely increases the probability of an effect obtaining. She argues that probabilistic causation is well established in physics now, for example radioactive decay. But then the question becomes 'how is the intention under control?' If 'I prefer x' only raises the probability of 'I intend x' and does not necessitate it then control is lacking. Therefore Ekstrom locates the indeterminism prior to preference formation during deliberation. According to Ekstrom deliberation can issue in intentions or preferences. Preferences are similar to Frankfurt's second order desires but are also formed in a search for what is good and are thus connected to the agent's conception of the good. Acting on a desire is not necessarily free, for example in the case of addiction,[147] however acting on preference is free, since it is authored by the agent. However if preference formation is deterministic (i.e. the function of past conditions and deterministic causal laws) then the past becomes the author, and control is lost. What we need is 'undefeated authorisation' which is a preference that is not coerced and is not a product of the past through deterministic causation. Ekstrom (1993: 608) describes an authorised preference

[146] As it is for those volitionists who base control on some primitive mental act.
[147] Although there is reason to doubt Ekstrom is right about this, and that addicts do 'not have a choice' – see Pickard (2013).

as one that coheres with an agent's other preferences. The set of preferences that an agent would defend are indicative of the agent's true self, or character. Ekstrom notes that deliberation resolution does still look arbitrary, and she hopes to alleviate this by explaining her conception of an agent. An agent is marked by a certain character and the power to change that character – reform. Character is the input(s) that cause but do not determine the outcome of deliberation. Therefore whatever the outcome, it's source will be the agent's evaluative faculty (and the idea appears to be that it cannot be traced back any further). Determinism would threaten this model if character formation was determined by the past, but according to Ekstrom this is not the case because of the indeterminism inherent in character reform.

The key issue for Ekstrom is that her account does not sufficiently ground control, at least, not the sort of control required for freedom of choice. We require a real ability to control which of the forking paths we follow, rather than the counterfactual possibility that other forks could have been open if the 'trunk' preceding it had been different. On Ekstrom's account the forks appear at the point of deliberation prior to preference formation. She claims that the agent is in control of which alternative obtains because it is the agent's own character that guides the 'choice' between forks. However I do not think that the character's involvement in the process is enough to ensure control. The character raises and lowers the probabilities of various forks obtaining through a complex coherence relation that each fork shares with the agent's set of preferences. So there is a sense in which the agent is in control of the probabilistic lie of the land prior to the outcome of deliberation. The agent directly affects the likelihood of any fork obtaining, and they do so in virtue of who they are, not in virtue of randomness. However the luck objection can still be made because the final outcome of deliberation rests with the indeterminism, and not with the character of the agent. By analogy: although the character may load the dice, for example by adding two dots to the 'one' to make it a three, when it comes to decision time the dice are rolled and could still fail to land in a way that secures the most probable outcome. There is nothing the agent can do to secure one particular outcome. Ekstrom has the agent as controlling the outcome of deliberation because the past and causal laws do not determine the outcome. The agent's character plays a controlling role in limiting which forks could obtain (although remember that it is a limiting of the probability of an outcome obtaining rather than limiting the possibility). However the agent has no control whatsoever concerning how these probabilities finally play out. The final say, the up-to-us-ness, is in the hands of luck (or indeterminism).

The role of indeterministic coherence evaluation threatens the agent's ownership of their character creation, so that not only are current preference

evaluations susceptible to luck, but the entire agent is. Remember that character is defined as the set of an agent's preferences and presumably the relations that these share. Her account of an agent is that it is a being that possesses character and the power to reform that character. So if each preference that is introduced to this set arises through the same indeterministic process Ekstrom has described, then the contents of this set cannot be determined by the past: an autonomy from the past has been secured. However we might doubt that the agent's true self can rightly be identified as character when, as already objected above, the final say about the contents of this set rests with randomness rather than the self. Ekstrom's proposal is complex and subtle, but ultimately fails to avoid the luck objection. There are many other varieties of event-causal free will, but we will not need to look at any more. By considering how the luck objection is fatal to these two theories, we have gone far enough to see why it is such a perennial problem for incompatibilists. These two theories demonstrate a general approach to the problem, and despite variations on their theme, the commitment to event-causal control and indeterminism will continue to raise the problem of luck.

Abandoning a causal analysis

There is a general problem with any approach to up-to-us-ness that attempts to explain agential control in terms of causation. We want agents to have control over which fork in the path they take, or which action they perform. Up-to-us-ness captures this idea of control as the agent being the one who determines which fork is taken, rather than it being something that happens to them. Causation looks like it will be an apt tool for explaining the exercise of control, because causes work by determining that one thing is possible (or certain) and making alternatives impossible.[148] So, event-causal theories attempt to explain what sort of agent-involving state or event must be the cause of an action, to satisfy that action being up to the agent. The 'right' chain of causes leading to an action enables the agent to exercise control over their action. But once this causal model of agential control has been adopted there is a problem as soon as indeterminism is introduced. Since control is constituted by an appropriate causal chain, a break in this chain would constitute a breakdown of control. But indeterminism just is the introduction of a break in this casual chain. Indeterminism means that whatever comes before cannot determine what comes afterwards. Now, if this indeterminism came before the agent's time of deliberation, or perhaps even before the agent was born, then we would not worry about this instance of indeterminism affecting their ability to control their action – as indeed compatibilists are wont to not worry about the consequence argument. This is

[148] Pink makes this point (2004:115).

because an appropriate causal chain could still obtain in the agent. But placing the indeterminism there means it cannot be something any state of the agent now has any influence over, or can interact with. It does not help us provide the agent with alternative possibilities they could influence with freedom of choice. So event-causalists must place the indeterminism somewhere within the causal chain we would identify as 'action producing'. But doing this undercuts their ability to tell a causal story about the agent's control. If control obtains when the right sort of cause determines what happens, then the absence of this cause doing the determining signals that control does not obtain. When indeterminism is introduced, luck, or chance, or randomness, is introduced as an additional determinant of outcomes (or so we have argued).[149] But this indeterminism cannot be controlled, because in order to be controlled it would have to be causally determined, in which case, it would no longer be an instance of indeterminism. So event-causal theories seem doomed to fail. The marriage of a causal theory of control, and indeterminism, is destined to lead to the luck objection.

Pink (2004) & (2011) has noted that we already have reason to be suspicious that causation will be the right tool for analysing free control. Causation does not look like the sort of power that will be able to explain freedom because they are fundamentally different types of power. Causation is a one-way power; a cause can only produce the effect it is a cause for. Causes do not have control over the kinds of effects they produce. Freedom is however a two-way power (something that Kane talks about at length but does not connect with the difference from causal power).[150] Freedom is a power to do *A* or *B*. For example, if I throw a brick at the window the glass will break, assuming the brick is thrown with sufficient force. When the brick hits the glass, given the properties that the brick possesses as a cause and the properties the glass possesses to be operated upon, there is nothing that can happen other than the glass breaking. Something may prevent the glass breaking, but that something would be another cause operating on the brick and glass considered as a system. My decision to throw the brick is different; I can decide to throw the brick, or not to throw it. If decision is an exercise of freedom of choice then it is a two-way power. I have control over which decision will obtain, and which action I will perform, and self-determination means that the obtaining of one rather than the other is something that I deliberately do. The way I act is up-to-me. But my exercise of control does not entail that one result *must* obtain, or had to obtain. By exercising control, the result is the obtaining of

[149] We should remember that this is also the case for probabilistic indeterminism. The influence over probabilities can load the dice, but when they are rolled, the result will be subject to chance.

[150] Some people prefer the term multi-way power, as we need not be limited to only two options. This is right, but as long as we remember that we mean 'at least two' this term is not problematic.

one of the possibilities. But whatever mechanisms I possess that make it possible for me to control, and that are so constituted as to enable my controlling this action (or decision), do not make it the case that one possibility must obtain – a different possibility could have done. Causation determines, and thus controls, by making any alternatives impossible. However our common sense conception of freedom suggests that our exercise of freedom to *A* excludes randomness but still leaves it a possibility that I could do *B*.[151] Therefore freedom and causation exclude randomness, i.e. determine in a controlled way, in intrinsically different ways, which suggests that it would be a mistake to try to explain free control in terms of causal control.[152] So the power exercised by an agent in freedom of choice situations (which we can appropriately call the power to choose), is unlike the power of causation. Because they have fundamentally different properties, it is a mistake to try to reduce the former to the latter. So we have a number of reasons to suppose that we are not going to be able to explain how action is controlled by utilising causation.[153]

Does agent-causation help?

But what about agent-causation?[154] It is frequently charged that agent-causation does not resolve the problem, but merely labels, or renames it.[155] This is not obviously true of all agent-causal theories, but it is worth considering why it could be a problem.[156] If an event-causal theory of control is adopted then a regress of explanatory events is generated. Events control their effects by being sufficient for their occurrence, but each event identified as relevant to the exercise of agency must be random or controlled. So to preserve control we postulate a prior sufficient cause, which is another event, to protect the theory from

[151] For a good discussion of whether this leaves what is done as inexplicable see Ginet (1995), and other essays in O'Connor (1995).
[152] This is noted by Pink (2004:115).
[153] Note also that this insight explains why libertarians run into problems when they are asked for contrastive explanations of an agent's acting as they do. Libertarians should not be drawn into providing a contrastive explanation, for it begs the question against a non-causal account of control. See Mele (2000) for an instance of this question begging.
[154] Sometimes agent-causation is charged with being metaphysically weird, because it is thought that it tries to ground the agent-causal power beyond, or above, the natural causal order. However this objection need not concern us, since there are versions of the theory that allay these concerns, and in the context of a project concerning theism and heaven, we have already admitted what these objectors conception of the natural order would count as metaphysically weird anyway. See Clarke (2000).
[155] This can be an issue for volitionists, where it can sometimes be charged that the will is personified. A theme we see from antiquity through to scholasticism, see Steel (2004) and Pink & Stone (2004). Note also that agent-causation is not a label applied only to incompatibilist accounts, see Markosian (1999).
[156] For example Lowe (2001) identifies important agent involving events involving intention, but denies that a causal explanation of these events is possible. See also Pink (2004)

randomness. But then we ask again whether this event is random or controlled (it could of course be determined and not random, but out of the agent's control, however we are not interested in this possibility). Agent-causation identifies some event in the production of action, often decision (though some theories identify the action itself, and many conceive of the decision as itself an action), and to protect from randomness puts the agent in a direct causal relation with that event. The claim is then that since the agent is not an event, the regress is terminated. The agent stands in the relation of substance cause to the event of the action or decision. Now this may prima facie seem appealing. After all we are looking for self-determination, so a theory that proposes that the controlling factor is the agent has clearly grasped something of what we are looking for.[157]

However labelling the need for a controlling factor as 'the agent' has done nothing to explain how this control is exercised. Objectors want to know what it is about the agent, this substance, that explains the agent acting, especially when it comes to explaining why the agent acted when they did. After all an agent can be in possession of reasons and still not act, so what is it that takes us from an agent primed for action, to an agent performing an action? How are agents 'moved' into performance of actions? Because we are still analysing control in terms of causation, there must be something about the agent, qua substance, which makes a difference to the effect it produces. Candidates for this difference could be changes in the properties of the substance, or in relations the substance has. But these, objectors will argue, are events. If it is denied that there is any property that makes a difference, objectors will not be happy to call it an instance of causation, or will again call it chancy. Pink (204:108) notes that we need to be careful not to assume that the only way something can be determined is causally. We have seen that there is reason to be concerned about whether causation is the right kind of power to associate with freedom. However we do want to say that agents have the power to determine how they act. So, should we try to analyse freedom of choice non-causally? This power of choice in acting does seem to be a different sort of power to causation – so yes, the non-causal approach does look like our only possibility for analysis of freedom of choice. I suspect that some agent-casualists have intuitions that would lead them to agree with much of what we have discussed about the difference between agential self-determination and normal event-causation.[158] But they would say that causation is still an appropriate term, as long as we distinguish 'agent-causation' from other types of causation. As long as the dispute is terminological, it does not matter, but in order to bring these

[157] See Chisholm (1964), O'Connor (2000), and Clarke (1993) & (2003) for important developments of the agent-causal approach.
[158] There seems to be a significant overlap between these views, although there can be differences.

differences to the foreground, we should say that freedom of choice is not a causal power, but it is still a power exercised by agents to determine how they act.

Non-causalism, agents, and whether actions are events

An immediate challenge to conceiving of freedom of choice non-causally is that a causal explanation can, apparently, still be asked for. However the agent's action is controlled and determined, if it is an event, then it does have a cause, since every event has a cause. The principle that every event has a cause is rather underdeveloped stated in this form, perhaps even naïve, but it does plausibly point in the direction of something insightful. Even if some events exhibit some indeterminism, such as radioactive decay, there are still casually relevant factors pertinent to the opportunity for an event's occurrence. But we do not need to argue that any such principle is false. Having admitted that there is some intuition here that may be troubling, it points us in the direction of examining the status of actions as events. If actions are not events, then the objection is averted, however the principle is construed.

If actions are identified with events, then the difficult question is, which ones? Suppose that I laid out the rake, lured my neighbour into my garden and then distracted him at the crucial moment, leading to his tripping and impaling himself on the rake. Is the event of his death something that is a doing of mine? What about the subsequent repossession of his house, and the depression suffered by his children? At some point it becomes inappropriate to call these events a doing of mine, though drawing a clear line here is difficult.[159] Some have suggested identifying actions with bodily movements. However this suggestion runs into problems since some of what we consider actions do not involve movements at all, for example if I do not want to be disturbed I can refrain from answering the telephone (Hornsby, 2004).[160] In this case my refraining seems to be something I can decide to do, even intentionally, and is a plausible candidate for an action. But it does not involve my moving my body in any way. Indeed, some of the things I do are to think, to calculate, to watch, to enjoy listening to, and so on – so called mental actions. So the physical location of the event does not seem to help us identify what is an action. As well as a problem with identifying events proceeding from an action, or the obvious point at which our consideration of the agent doing something begins, there is a regress problem. If an action is an event that is caused by the agent, then we will next ask whether the causing of this event is itself

[159] The problem of regress in identifying basic actions, and individuating actions is discussed by Goldman (1970), and explanations that do not rely on causation, and thus avoid the problem, begin with Stoutland (1968) and Chisholm (1964)
[160] See also Bach (2010), McGrath (2005), Thomson (2003) and Alvarez (2001).

another action (see Alvarez & Hyman 1998:222).[161] If it is, then a regressing chain of events is established whenever an agent acts. It is difficult to see how we might terminate this regress, or avoid it altogether, once we have claimed that agents cause their actions, and that actions are events.

We can utilise a distinction made by Hornsby (1980) between an agent's arm moving, and an agents moving of her arm.[162] The event of my arm moving may not be an action, for example if you moved it while I was asleep it would not be an action. But my moving of my arm is the sort of thing that can be an action. Actions have a possessive description, but the possessive always qualifies the subject of the verb which describes the act, who is the agent, and not a part of their body, a place, state, or time. So 'my raising of my arm' and 'Richard's raising of an arm' are both statements referring to an action. But 'my arm's rising' and 'Richard's arm going up' are not; in the first the possessive qualifies the object of the verb, and in the second there is no subject. Now, if actions are events, then should we identify the event of my arm moving as being the same event as that of my moving my arm? This move is frequently made. But it appears not to be the right move, because my moving of my arm is the cause of may arm moving. Events cannot cause themselves, so there must be two distinct events here, or my moving of my arm is not an event. As Hyman puts it 'an action – the causing of a change – cannot be identified with the change caused, because causation is a genuine relation' (2006:156). But if my causing my arm to move is a distinct event, then we face a regress again. So, we are faced with the alternative proposal, that my action of moving my arm is not an event. The action of my moving my arm is not the motion of the arm itself, so what is it?

Normally the event of someone doing something is an event of their bringing something about. Hornsby (2004) uses the example of a tea-drinker who puts her cup on the table, where the result of the event of putting it on the table, is that the cup is now on the table. If I move my arm, the result is that my arm is raised. By acting I cause something to happen, some change to occur. My action constitutes my changing something in the world, making some state of affairs obtain, but it is not the event of my causing this change. Alvarez and Hyman make the point clearly saying,

> a sentence which reports an action entails that an agent and an event stand in the relation 'is the causer of'… to act is to exercise a casual power – to cause,

[161] We will not consider the possibility of basic actions being caused by some prior volition, since ths does not appear to terminate the explanatory regress when causation is the tool of analysis, but for an attempt at this see McCann (1975 and 1998). The idea also appears in Reid (1843).

[162] This is normally noted by subscripts where to move_T is a transitive use of the verb, and to move_I is intransitive. We do not need to examine this detail, but it will be a useful reminder to those used to these terms.

> bring about or effect an event. But the exercise of a causal power is neither an event, nor the relation between agent and event that it entails. An action … is a causing of an event by an agent (1998: 233).

So events relate to actions as effects, as Hornsby has it, 'what agents cause, then, are not the events that are their actions, but the effects or results in terms of which their actions may be described' (2004:18).[163] So there is a sense in which 'agents as causes' is relevant to agency, but it is not that of agent's causing their actions. An action is not the effect of the agent, but an action constitutes the agent's causing some state of affairs to obtain.[164] Another way of noting this, though slightly awkward, is that actions are not something done by the agent; they are a *doing* of something by the agent.[165]

This helps us to understand why it is that the exercise of a non-causal power by an agent to act one way rather than another counts as controlled rather than random. When an agent exercises their freedom of choice, and acts a certain way, they do not control their action by doing something else, their action just is their controlling how they act, what results they bring about, and what ends they direct their agency towards. An agent's action is not the effect of their control, it constitutes their controlling how they act.[166] Remember that we have argued that performing an action requires an agent to possess motivation towards an end, or ends, that they take their action to be a means to satisfying. When an agent acts they make one (or more) of these ends ones that they direct themselves toward. The agent acts on a motivation they possess, and their action constitutes acting for that reason (in order to satisfy that desire) because it is an application of their capacity for practical rationality to how they act.[167] Agents are not caused to act by their reasons and desires, otherwise they would not be exercising control – they would be being controlled.[168] Instead agents possess a multitude of reasons and desires, and they exercise their freedom of choice over how, and when, they act on those reasons, or satisfy those desires, by acting – which is intrinsically doing something for a reason.[169] Performing an action constitutes my ability not just to move my body, but to direct my self towards certain goals – and directing myself

---

[163] See also Stout (2010).
[164] See Steward (2000) on this issue from the perspective of whether a causing is an event that could be in an agent.
[165] See also Ruben (2003) for criticisms of viewing actions as events, and also the place of this assumption in the causal theory of action.
[166] Pink (2011) emphasizes this relation in terms of constitution. We could also articulate this theory of action by distinguishing it from volition based theories, which are closely related, but still causal in character, see Pink (2008).
[167] Remembering that we have not said anything about the need for this capacity to involve conscious deliberation, though, in paradigm cases we are often interested in, it does feature prominently.
[168] We will discuss Davidson's objection to this in a moment.
[169] This point is also made by Alvarez (2009).

at a goal involves orienting my actions towards the achievement of some effect (end) which my action results in, in order to satisfy some desire I possess. The connection between my desires and reasons, and my actions, is not a happy accident of chance, nor is it that that of the former causing the latter. My action *is* the exercise of my capacity to possess desires and reasons, and to act upon some of them. So this capacity for choice is wielded not by my reasons, or desires, or some mental state, but is wielded by the self, the agent.[170] What the agent is, is a complex question, but in our context the agent is at least the instantiation in a substance of certain powers, including the power of choice, and the power of causation, along with the faculties requisite for the exercise of those powers – so the ability to form and to hold beliefs and desires, for practical rationality (deliberation, intention, means-end reasoning, etc.), perceptual sensitivity to the world around them, and the ability to conceptualise the good and make evaluative judgements.[171]

This commits us to the view that agents are not reducible to a collection of mental states (or their physical realisers). It is a benefit that we are committed to the position that it is an agent that performs an action, and not a state, or part, of that agent. As Velleman (1992) has argued, an agent is ineliminable from our description of what happens when someone acts. Although I do not want to argue at length against the causal theory of action, Velleman's comments are apposite.[172] The standard Davidsonian causal theory describes an action as an agent's desire for a certain end, and their belief that a certain event would be likely to further that end, jointly producing a bodily movement that will contribute to, or constitute, that end's obtaining. Velleman complains that,

> In a full-blooded action, an intention is formed by the agent himself, not by his reasons for acting. Reasons affect his intention by influencing him to form it, but they thus affect his intention by affecting him first. And the agent then moves his limbs in execution of his intention; his intention doesn't move his limbs by itself (1992: 462).

So the initial suggestion, that actions are things that agents do, requires some qualification. The 'things done' are the effects of the action that allow us to identify and describe the action, or the event (act), which constitutes performance of the action. But the action is not the event itself, but the doing of the act by the agent. Actions do not describe events, processes or states, but are about how agents interact with the world to produce events and processes, and maintain

[170] So there is a close connection between non-causalism and agent-causalism.
[171] There is, of course, a lot more to be said about this question that we do not have space to tackle, but see the later discussion of character especially.
[172] Taking on the causal theory of action in detail, would be a project in its own right, remember that we will only need to say enough to motivate the claim that freedom of choice is coherent.

states. So the action of an agent is not reducible to the event of the agent performing something.[173]

Non-causalism versus apparent causal explanations of action

However we have not finished dealing with our original objection, because if there is an event of the agent acting in a certain way read in terms of 'an event of such and such a result being brought about', then we can still ask about the causes of this event. And if a causal explanation can be offered, then this would seem to undermine the non-causal strategy we have developed.[174] However our comments about the centrality of the agent will help us to understand the place of any such causal explanations of agent-involving events like this. We should expect an examination of action explanation to help us to identify and understand something about agents, because if actions are the doings of agents, then agents should figure in the explanans. If agents were not a part of the explanans then it is difficult to see how we could possibly establish that an action was a doing by an agent. This was Velleman's point above (which is really more about action explanation, than action); it is also made by Hornsby in critiquing the causal theory of action's explanation of actions as being caused by desires.

> For even where there is an event of the agent's doing something, its occurrence is surely not what gets explained. An action-explanation tells one about the agent: one learns something about her that makes it understandable that she should have done what she did (2004: 8).

We need to take care to distinguish several types of explanation of action (or agent involving events).[175] Sandis (2012a:333ff.) has distinguished three senses in which action explanation is proffered,[176]

> i) Why A's body moved
>
> ii) Why A's action of moving her body occurred
>
> iii) Why A moved her body

Sandis explains that,

> In (i) we are searching for the cause of the event that was the movement of A's body. What is sought in (ii), by contrast, is the cause of the (related) event of A's moving her body to occur. Finally, to provide an explanation of (iii) is to explain […] why A performed the action of moving her body.

---

[173] See also Wiggins (1998) on Hume and causal/contrastive objections.
[174] For a recent collection on the issue of whether reasons are causes, and the causal theory of action, see D'Oro and Sandis (2013).
[175] For an overview of the need for distinction see Stout (2005), and also Goetz (2008).
[176] See also Sandis (2012b) for a fuller discussion of the effects of conflating these senses of explanation on the development of action theory.

We have already argued that the movement of an agent's body should not be confused with the agent's moving of their body. So it is no surprise that if these are distinct things, even in different categories, since actions are not events, they will have different types of explanation.

Consider the example of my raising my arm to gesture to a friend that I see walking down the road towards me. My action is my raising of my arm – my causing of my arm having been raised. However there is also the event of my arm rising, the actual bodily movement, and there is also the event of my acting in such a way as to raise my arm. The explanation of my action *itself* concerns iii), and explanations of 'causings' by agents must be teleological in nature, that is, explanations of why an agent acts as they do must refer to goals that the agent adopts – the ends at which their action aims.[177] Since actions are the exercise of practical rationality in an action constitutive way, and since actions intrinsically involve goal-orientation by the agent, the ends at which an agent aims by so acting must figure in the explanation. Alvarez (2013:147) considers the example of an agent watering the plants at dusk, since watering them in the middle of the day would be ineffective because of evaporation. The agent has a reason for watering the plants at dusk – some fact about evaporation, the loss of water, and the need for plants to have time to absorb water to keep healthy.[178] But the reason does not explain the agent's action, because they need to be aware of this fact for it to have anything to do with their action. But even if the agent is aware of this fact, still it does not explain their action, because without the agent possessing some goal at which their action aims there would be no connection between the beliefs (or reasons) the agent has, and the kind of action they perform. The fact about evaporation only explains the agent's watering the plants at dusk on the presupposition that the agent has a goal of keeping the plants healthy. If the agent has the goal of keeping the plants healthy, or economising on water, then the reason is relevant to how they act. But the possession of desires that could motivate an agent towards some end or goal is not sufficient for the agent actually being oriented towards a goal. Agents orient themselves at goals *by* acting – which can include actions such as making decisions, forming intentions, making practical plans, making evaluative judgements, and so on. Therefore the event of

[177] See Sehon (1997) & (2005), and McLaughlin (2013) for more on why purposive explanations are non-causal.

[178] I am not going to discuss internalism/externalism about reasons, as this would be a considerable undertaking at this point. For the record I agree with Dancy (2002) that reasons are states of affairs, and not mental states. However, were someone to be committed otherwise I think that there is some scope for internalists to construct workarounds to allow reasons to play the role in freedom of choice that they need to, even though this may not be the best solution. So, it is not essential for us to discuss here. For more discussion of the superiority of externalism, see Alvarez (2010), Dancy (1995), and also Stoutland (2001).

coming to realise that there is some reason, or the acquisition of a desire (in other words, events), cannot provide an explanation of why an agent acts.[179] So the explanation of my raising my arm could be something like: 'in order to greet my friend', it will give the goal at which my action aimed, a goal I adopted by so acting.[180]

However, another explanation that could be offered is that I saw my friend walking down the road. The event of my seeing, if this does explain my acting a certain way, would then be a casual explanation, which would be problematic. But this explanation is not really about why I acted a certain way, instead it is an explanation of why my action of raising my arm occurred, or to be more precise, why the event of my acting in such a way as to cause the result of a raised arm occurred. These questions correspond to Sandis' category ii). The key insight here is that although there is something pertinent to understanding the agent's action by considering the event of my friend walking down the road (or my seeing them doing so), it does not undercut what we have said about event-causal explanations destroying the ability of agents to possess freedom and control. This is because although there is an event of my friend walking down the road, whose casual influence is felt by me, the agent, and my faculties of perception, belief formation and so on, the occurrence of this event need in no way be seen as something that determines the occurrence of my action. This event does provide the opportunity for me to gesture to a friend, but it does not necessitate it.[181] One way to look at this is that the occurrence of this event is a structural cause that makes it possible for me to act on a motivation (perhaps I have been longing to run into a friend and greet them). Or we might say that the occurrence of this event introduces a reason that the agent could act on, but whether they will do so or not is not determined by the possession of this reason. However the introduction of an event-cause, whether we consider it in terms of being a structuring cause rather than a triggering cause of my action, or whether we just say that it is an indeterministic contributing cause, does not introduce any threat

[179] Sehon (1997), (2005) and (2010) argues in considerable detail about the irreducibility of teleological explanations. See also Schueler (2003).

[180] It may be that I saw my friend some way off, and did not raise my arm until they were close enough to see my face. If I decided I would raise my arm, then I adopted a goal at that point. But my decision, or formation of an intention, does not cause my later raising my arm – after all I could have failed to do so. But my decision does alter my motivational state, my awareness of certain reasons I possess, and the focus of my attention. In a way I orient myself at the goal of greeting my friend when I make the decision, but the tense of this statement matters. When I decide my goal is that 'I will greet my friend', and when I actually raise my arm my goal is 'to be greeting my friend'. These two are related. This is a relatively minor point, but worth consideration to avoid thinking that agents can only adopt goals by present bodily movements, or any such similar view. We will consider how our motivational state may influence how we act shortly.

[181] See especially Dretske (1988) and (2010) for more about structuring causes and action.

of randomness, because an agent's control is exercised and explained by their acting a certain way, and that fixes us squarely in the realm of iii). So the presence of a causally structured explanation of my action that cites an event is not in competition with our non-causal analysis of freedom of choice, as long as we distinguish explanations of my action, from explanations of the event of my acting.

The other sort of explanation of my action that could be offered might go something like this: 'my arm rose because a neural motor control system caused certain muscle groups to contract', or a variation on this theme. This is an explanation of why my arm went up, which corresponds to Sandis' i). The comments we made above are also pertinent to these sorts of explanation. As long as the neurophysiological event-causal chain leading to and including my arm going up is not deterministic, then there is space for the agent to do otherwise. But, once again, this stipulation does not mean that we have introduced the threat of randomness again. For an agent's control is exercised in the constitutive relation between his capacity for practical rationality and his action. In fact, if this chain of events is a description in finer detail of what it is for an agent's arm to go up, or an agent to raise their arm, then these neurophysiological details will be part of the result, caused by the agent when they act a certain way. Alvarez puts it this way,

> And it seems plausible to suggest that these motions [my arm going up] are caused by neural states and events and, more generally, by a complex of neurophysiological occurrences and conditions. However, these motions are not themselves actions but the results of actions: they are (among) the things we make happen when we act (2013:153).

But there is a prevalent concern that our neurophysiology is a physical system, developing deterministically according to well-defined natural laws, and that this would prohibit any agent from having the power to make anything happen otherwise. So although the admission of these causal explanations having a place in explaining agency does not prohibit control, it may prohibit freedom of choice, because there could not exist alternative possibilities. Although there are sceptics about it's plausibility, we noted that Kane has suggested that quantum indeterminacy could be present in neurobiology, and could be amplified through chaos inherent in the complexity of the system to provide alternative possibilities. However, even if this is so, it is difficult to see how a quantum fluctuation could be something that is a result of an agent's power to act. It may be important for helping us see why there is sufficient causal slack at the physical level to allow for other forms of causation, without breaking physical laws.[182] There are other

[182] See Ellis (2009).

prominent proposals. O'Connor (2000) has championed the idea that our complex neurophysiology may exhibit a complexity sufficient for the emergence of non-physical properties, and associated powers and relations. Emergent dualism avoids the 'metaphysical wierdness' of Cartesian dualism by grounding the existence of emergent properties in the natural realm we are familiar with.[183] It is also independently motivated since emergence is used to explain the behaviour of biological systems other than in the context of action. A second possible avenue of explanation is that our neurophysiology allows for consciousness (even if physically realised) to play the role of bringing together multiple parallel strands of possible influence, and resolving them – the global workspace theory of consciousness.[184] Although I suspect that some combination of emergence and a global workspace theory may get us what we need here, we do not need to concern ourselves with the details. Both of these suggestions are forms of top-down causation, which make room for the sorts of alternative possibilities that could conceivably be a result of an agent's action. If it is possible that there is an answer to this question of how room can be made for alternative possibilities, even if the details are yet unknown, then the objection is not fatal, and the prospects seem promising, and there is no reason to conclude that freedom of choice is physically impossible.[185]

Davidson's challenge, and objections to the causal theory of action

The non-causal analysis of freedom of choice has survived the challenges presented thus far, however two important issues remain that deserve special attention. The first of these is Davidson's challenge, which concerns what it is for an agent to act for a reason. The second issue is why, given that agents may do all sorts of things given the plethora of reasons and motivations available to them, agents persist in so often acting according to their strongest desires or reasons. Davidson motivates the move from a non-causal to a causal theory of action by challenging non-causalists to explain how we can analyse an agent's acting for one reason rather than another in the presence of multiple reasons for action. Davidson challenged the non-causal norm by noting that a 'person can have a reason for an action, and perform the action, and yet this reason not be the reason why he did it' (1963). So in virtue of what is it true that an agent acted for a

[183] See Clarke (2000) on worries about metaphysical 'weirdness', or extravagance.
[184] See Baumeister, Mele & Vohs, (2010), and Levy (2014), for a discussion of this theory. Although I have used the term 'parallel' the way I conceive of this is more akin to a superposition of cognitive and evaluative states, which the agent has the power to resolve one way or another. But for those unfamiliar with the metaphysics of superposition, parallel, is the next best term, and not much hangs on the difference.
[185] For more on these issues, and the prospects for a solution see Gibb, et al. (2013), O'Connor (2000), Murphy & Brown (2009), and Murphy et al. (2009).

reason? We can frame the challenge clearly by asking when an agent has more than one reason for A-ing, φ and ψ, what makes it true that they acted only for φ and not for ψ? Davidson responded that the reason that causes my action identifies which reason I act for (1963). For example, suppose I am the judge of a competition. Your being my close friend gives me a reason to award the prize to you, but also, and in a judgement quite independent of your being my friend, your entry to the competition is the one I regard as the best, which furnishes me with another reason to award you the prize. If I award you the prize, in virtue of what is it true that I awarded you the prize for the reason that your entry was the best, and not, unfairly, because you are my friend? For Davidson, the answer can only be that the former reason caused my action, whereas the latter did not causally contribute to the production of the action.

Davidson's challenge motivated the development of a causal theory of action that had the resources to explain action in causal terms sufficient to answer this question. The causal theory of action, or varieties of it, dominate the philosophy of action, and although the causal theory of action we will examine is not Davidson's it is Davidsonian, and represents a core consensus in broad brush terms.[186] The causal theory of action claims that we can identify a bodily movement as an agent's action by the causal antecedents of that bodily movement. As current orthodoxy has it, when an agent has a desire for *X* and a belief that their *A*-ing would be likely to produce *X*, this belief/desire pair are jointly sufficient to cause that agent to *A*, where *A* is the bodily movement that constitutes their action.

We have already given some discussion to what it is for an agent to act for a reason, but it would be good to highlight these claims again in response to Davidson's challenge. We argued that when an agent acts they do something that is intrinsically goal-directed, and in order to be an action, that goal must be an end that the agent desires, and the agent takes their action to be a likely means of achieving that end. The possession of a desire for an end, and the prospect of the means of satisfying that desire (or achieving that end) give the agent a reason to act. Their reason for acting is 'to satisfy this desired end', which in our example was one of either 'to please my friend', or 'to reward the best entry'. Davidson's challenge effectively begs the question against the non-causal account by pointing to goal-direction, which is acting for something, and then pointing to the action itself, in terms of the bodily movement, and asking what connects the two together

[186] See for example Mele (1992 and 2003), Davis (2005). Another interesting issue for the causal theory of action, is whether the connection between beliefs and desires is necessary, but we will not explore this because of space, though the issue may run parallel to some of our previous discussion. See Melden (1961), Stoutland (2001) and Dancy (2004), a brief defence on behalf of the causalists in given in Mele (2010).

in the right way. However for the non-causalist the bodily movement that is part of the action, minus the goal-directedness, is just not an action anymore. So the challenge presents somewhat of a problem, if it is accepted as legitimate then the non-causalist has agreed to answer a question about what makes a certain bodily movement acting for a reason, when they do not even think that the movement in question is acting any more. It is therefore no wonder that once this concession has been made, answers seem worryingly incoherent. What the non-causalist should do, is to object that the challenge is not legitimate as the question itself belies a fundamental misunderstanding about what actions are.

Actions are all goal-directed, they all aim at an end, and the power agents exercise in freedom of choice is to choose which ends they act for (which can be more than one). Performing an action constitutes the agent's ability not just to move their body, but to direct themself towards certain goals – and directing themself at a goal involves orienting their actions towards the achievement of some effect (end) which their action results in, in order to satisfy some desire they possess. The agent's action *is* the exercise of their capacity to act for a reason. So if an agent possesses two reasons, one to make their friend happy, and the other to reward the best competition entry, the question of which reason they acted for is settled by looking at the intrinsic property of the action itself. Often agents will be conscious of these properties introspectively, but if not then we may not have epistemic access to the answer.[187] So the answer to 'what makes it the case that the agent acts for one reason rather than another' is 'the way that the agent acts'. If the agent acts in order to make their friend happy, this identifies acting for this reason, and if the agent acts to reward the best entry, the identifies this other reason. Davidson takes it that the actions in both of these cases are identical, and so the only differentiation between them has to be in terms of their causes. However, because they are inherently goal-directed, they are not the same action, and we can differentiate between them by considering the properties of the actions themselves.

We can do more to deal with Davidson's challenge however, because there are reasons to doubt that the causal theory of action can itself deal effectively with explaining under what circumstances an agent has acted for a reason. These objections give us reason to doubt the causal theory, weakening any qualms about the non-causal analysis. Firstly, we should consider how well the causal theory can answer its own challenge, or rather, some problems that arise from its answer. In

[187] It is worth noting that if, as Alvarez suggested earlier, neurophysiological events are ones we cause by acting, because they are among the results of what we do, then there may be clues as to the intrinsic properties of an action here (especially when the result is not just the final state, like an arm raised, but a process) which would be discovered empirically even though they are not introspectively accessible by an agent. This may turn out not to be the case, but we should not rule out the possibility at this time.

order to examine the causal answer we need some idea of what these reasons are that are supposed to be causing bodily movements. This is not an uncontroversial topic, but thankfully a sketch will suffice us. Reasons play a role in practical reasoning, which is teleological in nature. That is to say that it takes a means-end form. Reasons motivate agents to perform actions by rationalising the act, *A*-ing, as a means to achieve a desired end, *E*. For most causalists (who are normally paid up members to the project of naturalising action explanation) the cause of the action is taken to be a mental state (or event, depending on your ontology) that can stand in the relation to the bodily movement as efficient cause, therefore motivation is given a causal reduction. The details of neurophysiological mechanisms do not matter to us.[188] Whatever they are, they are physical realisers[189] of pro-attitudes towards *A*-ing; that is, agents with reasons for *A*-ing possess a pro-attitude, or desire, towards the end at which *A*-ing aims, and this desire motivates the agent to *A*. Although my beliefs and abilities are causally relevant to what actions I perform, it is common to see a focus on the motivational role of desires in causal theories. Causalists, like Mele (2003), take an agent's desiring *E* to be the reason for their *A*-ing.[190] My desiring *E* motivates me to act to satisfy that desire and is the reason for my *A*-ing.[191] So my motivational state causally explains why I act, and also reveals my reason for acting (hence Davidson's challenge is answered).

To return to our example, we recognised two reasons presenting themselves to my practical reasoning: your being my friend, and your entry to the competition being the best. So I possess two mental states, a state of desiring to celebrate our friendship, and a state of desiring to honour the best entry to the competition. Note that I have substituted plausible ends to be the objects of my desiring, since under the above analysis it is ends that are desired, however other analyses of the objects of desire are available, but I do not think it matters to my point here. Now, I award you the prize, and you want to know the reason why I did so. The reason will be either my desiring to celebrate our friendship, or my desiring to honour the best entry to the competition. Somehow, we are able to ascertain that the latter reason, i.e. mental state, was causally responsible for my

[188] Though, some have argued that the nature of the ontology of mental states could provide an objection to the causal theory, see Sehon (2000), or for broader issues Heil & Mele (1995).

[189] The relation of realiser is also controversial, but alternatives could be substituted here if desired.

[190] These are so called motivating reasons, see Everson (2010) for a discussion of this conception of a reason, and Alvarez (2010) for a more extended critique.

[191] It is worth pointing out that I think causalists are guilty of equivocating between 'the reason for my *A*-ing' and 'the reason for which I *A*', see Alvarez (2010) and Sandis (2012), and that this may lie at the root of where the causal story is going wrong. But my task here is not to argue that point, only point out a dilemma for the position as it is commonly construed, so I let sleeping dogs lie.

action.[192] You are understandably relieved; I was not motivated by your being my friend, but by the quality of your work.[193] Remember that motivation is given a causal reduction under the causal theory of action, and that although desire has the pre-eminence in causal explanation, these factors are intimately causally relevant to the possession of the relevant states. Another way of talking about this would be that our friendship was not causally responsible for the award, but the quality of your work was.

Let us think again about the role of reasons in motivating action in situations where the agent possesses multiple reasons for action. Since reasons are motivationally relevant only when they are causally relevant, any reasons that are causally superfluous to the actual efficient causing of my action do not play a role in motivating my action. If I act for a reason, in the presence of other reasons, then although the other reasons may have a role in deliberation etc. they cannot motivate my action, nor even be something that I desire in acting, on pain of contradiction, since what motivates me is my desiring, and my desiring of something is my reason for acting. Now, this seems to be a strange result, and one that we should be loath to accept. Why should we not say, as I take common sense folk psychology to say, that I am motivated by all of the reasons that supply me with rational grounds for my action? I may feel bad about awarding you the prize for the sake of our friendship, but it certainly motivates me to ensure you receive the prize. Even if this sounds too objectionable, casting the objection in terms of desire may seem less so, since it is difficult to deny that in such a situation, when I award you the prize, I desire to celebrate our friendship *as well as* honour the best entry. If this were not the case then why would it be more special for me to be able to award the prize to a friend than to a stranger? If I did not possess a concurrent desire to celebrate our friendship then there would be no aspect the experience that could contain an element of that desire's satisfaction. Or consider the concern I might feel, as I award the prize, that I am doing so for the right reason. Why would I experience a prick in my conscience unless my desire to celebrate our friendship, and the concomitant reason, are not still a part of my psychological state as I do so?

It is clear that we can act in the presence of multiple reasons and desires. But the critical point is that in the presence of multiple reasons and desires, it can still

[192] Hutto (2013), in D'Oro & Sandis (eds.), notes that Davidson's proposal is not as helpful as it may appear in helping us to discover why agents act as they do, since it seems to require a sort of privileged access to internal causes with epistemic transparency, and quite why agents should possess this, especially in the face of so much empirical evidence that we do not, is far from clear.

[193] To be precise, I ought to say that it is the mental state realisers of my beliefs that you are my friend, and that your work is of a high quality, which do the causal and motivational work, but this précis is not damaging as long as this distinction is kept in mind.

be true of me that I act *for* only one reason. It still makes sense for me to say that I am pleased to give the award to a friend, but that I did not allow this concern to influence my decision. There may be some causalists who will deny the veracity of my introspective report that other reasons/desires did not in fact influence my decision. We should certainly not deny that there could be influences over our decisions and action that we are not aware of, and some of these influences are certainly of a kind that includes them as part of our practical rationality.[194] However the presence of non-conscious internal influences over our actions concerns the question of why an agent acts as they do, rather than the question of what is the reason(s) that an agent acts *for*. It is important to keep these questions distinct. Despite the influences over my decisions and actions that I might be unaware of it still seems a part of the way we ascribe agency to people that they can so utilise their capacities for practical rationality and agency that they can act *for* a reason, that is, they can take something to be an explicit goal, or end, at which their forming of an intention, or performance of an action aims.[195] It looks like the causal theory of action as we have examined it lacks the resources to make space for this feature of why agents act as they do.

Perhaps the causal theory of action position could be maintained and causalist could claim that the role of plural reasons in deliberation include a substantial enough role to account for the folk psychological observations. But this claim is not without problems, even if we allow that the influence of reasons in deliberation is dealt with sufficiently under causal theory of action. Why should reasons, which provide a pro-attitude, and hence motivation in deliberation, suddenly fail to motivate any more once I have decided how to act, or once I initiate my action? It cannot be that they are still reasons, pro-attitudes, and positively motivating, but are unused, or causally inert in action performance. Given the reductionist strategy of the typical causalist psychology once these reasons lose their pro-status, and any motivational strength, they cease to be reasons for action. How could my acting for one of my reasons suddenly render another reason irrelevant, non-motivational, or no longer reason-supplying? This would be a strange analysis of the role of reasons in action explanation. Agents normally reduce the influence of a reason by consideration of countervailing reasons, and not by discovering that it is not a reason at all once they act for a different reason. In fact, an agent finding themselves in a position where what they remember as a reason, φ, no longer supplies any rationalisation for action, or

[194] For example, see Levy 2011, 2012 & 2014 on the evidence for implicit attitudes.

[195] We could say something slightly stronger: that they make something a goal. But then the question arises whether they make something an explicit goal by forming a practical judgement prior to intention formation, or action, or whether their making it a goal is constituted by their so these. This is not important now, as in ither case, they would still be explicitly acting for a reason.

associated desire, is likely to report that φ never was a reason in the first place rather than its losing this status. While agents can be mistaken about reasons they could have acted on but did not (in hindsight) we certainly should not subsume all cases under this explanation.[196]

An alternative strategy could be to say that all of our reasons are causally efficacious in deliberation, but only the strongest reason is the efficient cause of my acting (I wanted to award you the prize because of the quality of your submission more than I wanted to award it because you are my friend). But this cannot be correct either. Under this proposal, we could reduce the standard causal response to 'an agent acts for a reason if one of their reasons is stronger than the others.' Two problems would then follow: agents would not be able to act *for* multiple reasons, and agents would not be able to act akratically for a weak reason. Agents would not be able to act for multiple reasons because it seems unlikely that two or more reasons would possess exactly the same motivational strength, and unless this could be ensured the stronger would become the efficient cause of the action, and the reason for which the agent acts (according to causal theory of action), leaving the other reason no longer a part of the explanation of the reason *for* which the agent acted (though it may, of course, explain why they acted as they did). It is a more natural causal picture, to allow each reason to supply causal influence in proportion to its strength, so that they are jointly causally responsible. But remember that this step back takes us to the beginning of our problem, explaining what makes it the case that an agent acts for one reason rather than others he possesses, so we would be no further ahead in answering Davidson's challenge. Agents would not be able to act akratically because doing so requires the ability to act for weak, and sometimes knowingly weak, reasons. We do not need to survey the many different situations that can be labelled akratic action. A case of an agent who acts for (on) a reason that they know to be one whose ends are of less value, and which they desire to avoid will suffice, such as succumbing to a minor temptation. The only way to avoid this problem would be to claim that akratic actions are still actions, but are not performed for the reason that causes them (which would be to reject causal theory of action), or they are not performed for a reason at all (which the causal theory of action could perhaps make space for, though standardly if a reason causes the action, it is the reason the agent acted for). In either case, it still seems that some agents can act akratically *for* a reason, which is itself akratic in nature.

[196] If there is some mechanism that alters our psychological states to 'purge' us of reasons that we do not act on in the scenarios imagined, then we might wonder why it does not operate at other times, for instance, when we act for more than one reason, that is, when we have more than one end at which our action aims. It is difficult to see how the causal theory of action could elevate acting for a reason to possess such strange causal powers while maintaining the causal reductionist strategy of analysing reasons as mental states.

This demonstrates that the standard causal response, a causal theory of action that is Davidsonian in tenor, is ill equipped to answer Davidson's challenge. Upon the adoption of causal theory of action we must either radically revise our folk psychological description of what acting for reasons is, or we are faced with considerable difficulties – features of acting for reasons that the causal theory of action cannot explain causally. So, the original motivation to adopt a causal approach to investigating action explanation, i.e. as a response to Davidson's challenge has been undercut, but there is one further prominent problem for causalists.

Deviant causal chains are a counter objection to the proposed causal theory of action.[197] According to the theory something is an action if it is a bodily movement that is caused by the right kinds of mental states, usually a desire and belief. However it is possible to construct cases in which this causal relation obtains, and yet the agent intuitively fails to act. Davidson (1973) introduces the example of a climber who wants to rid himself of the weight of his companion who is attached to him by a rope. Moreover the climber knows that if he loosens his grip slightly, then this will be effective in getting rid of the weight. But the climber does not decide to let go or loosen his grip, and forms no intention to do so. However the presence of this belief and desire so unnerves the climber that he inadvertently loosens his grip on the rope without meaning to, sending his companion to his doom. So, in this situation, the same belief and desire that would be appropriate causes of the bodily movement in the case of intentionally letting go of the rope are present, and they also cause the bodily movement in question. Yet intuitively, the agent has not acted, but is the victim of nerves. The challenge for causalists is to explain why this is not an action, and the response is usually some variation, albeit a very complex variation, upon the theme of 'causation did not operate in the right sort of way'.

Since the basic conditions have been met, what the casualist needs to do is to add a further condition that will explain why the deviant case is not an action, but leave non-deviant cases secure. But they may not invoke the concepts of agency or action in so doing, on pain of circularity. After forty years finding a non-contentious proposal for this further condition has proven decidedly difficult.[198] Steward (2012) notes that there is reason to be suspicious that this tactic will in principle not succeed. She notes that the general dialectic has been to note that the agent in the deviant case seems to lack control over his nervous twitch in which he loosens his grip. But what is control? It is a central concept in agency: that of

[197] For an overview see Stout (2005) and (2010). The size of the literature on deviance is considerable, and worthy of a project in itself. All we will do is identify a concern pertinent to motivating our rejection of the causal theory.
[198] For a discussion of some proposals, see Mele (1997)

some occurrence being under the agent's control. So control needs to be reduced into causal terms, and a general tactic is to add a sensitivity condition. This condition stipulates that if the circumstances were even slightly different, then the agent would have moved differently. That is to say, if the agent had intended differently, then they would have moved differently. In the deviant case slightly different intentions do not lead to different movements, so the sensitivity is lacking. For example, if the climber was considering his wanting to let go of the rope in a minute's time, this would still have unnerved him causing him to let go of the rope. However objectors to the causal theory are quick to object that this sensitivity condition can be met in situations in which agents fail to act, and that it therefore has not elucidated what control is. The usual cases involve omnipotent neurosurgeons who have disabled the agent's ability to move their body normally, but continuously intervene to move the agent's body in a way that is sensitive to their intentions. So the sensitivity condition is met, as is the causal condition, albeit via an intermediary, but intuitively the agent is not acting.

The next counter move is for the causalist to stipulate that the causal chain cannot involve another person's intentions, since such an interaction would constitute manipulation. But once again, counter-objections are made. What if the agent possesses a faulty faculty of motor control, but has be operated on to replace some nerve bundles with wires. So, the agent is now able to move their own body again according to their intentions. But, some of the wires have become severed, preventing the agent from exercising control. A helpful friend agrees to hold the wires together so the signal can pass down them, thus enabling the agent to control their body again – but this is only because the causal chain now involves the intention of the friend to hold the wires together. However, according to the condition added, this should not count as an instance of control by the agent. There is an intuitive observation that is relevant to resolving whether the agent is really in control, and that is that the agent is having his intentions enabled, and not blocked by the interaction of his friend. Because of its readily intuitive appeal we might fail to notice something Steward picks up on (2012:60). Saying that all is well as long as the process is one that belongs to the agent is tantamount to saying that all is well as long as the process is under the agent's control. But this is really just a clever way of saying that causation occurs in the right sort of way when it is under the agent's control, that is to say, the agent exercises their agential control over what happens. But this is clearly circular, since agential control is the very thing that the causalist needs to provide a reductive analysis of. So, the tactics employed in the dialectic seem prima facie to involve causalists in circularity, which will not surprise us. Our discussion of how to secure this sort of control took us directly away from a causal reduction. So deviant causal chains remain a

telling objection to any Davidsonian causal theory of action, and undercut the motivation to move away from a non-causal theory.

Explaining why agents tend to act on their stronger reasons

We have responded to Davidson's challenge by responding positively, and answering the question of what it is to act for a reason, and have also raised objections to motivate leaving the causal theory of action. Now we must consider the second outstanding challenge for non-causal theories: why is it that we so often act in accordance with our strongest reasons and motivations? As we have characterised matters agents can act on any of their reasons, as long as they possess the necessary prerequisites to be performing an action, it is an open possibility for them to perform an action, no matter how irrational (in the sense of silly, or imprudent) that action may be, or how weak the motivation or reason might be. On the positive side, this makes room for agent to be irrational, silly and acratic, which agents clearly sometimes are. However another incontrovertible observation about agents is that the very often do act sensibly, rationally and prudently. When agents deliberate about what to do, they do make choices that seem to align with the best, or strongest, reasons for action. On a causal theory this is easy to take some account of: stronger reasons and desires can exert a stronger causal influence on the outcome of deliberation, or action production. But for the non-causalist it is not obvious to see how this apparent influence can be accounted for, without relying on a causal conception of the influence of reasons and desire over outcomes.

What could it mean to say that a motivation or reason is stronger than another without appeal to causation? Although the causal approach is not without its puzzles,[199] there is something intuitive about causal talk here, indeed it is part of our everyday way of talking about our reasons and motivations. The idea that stronger reasons have a greater probability of being acted upon than weaker ones, because they exert a stronger causal influence over the direction of a deliberation, provides a simple causal theory based on probabilistic causation.[200] What we need is a non-causal analysis of this intuitive idea. Another way of thinking about this problem is that the various strengths of our desires and reasons can be thought of as having something to do with our character, or dispositions. Now, it is highly plausible that a model of agency should allow character to have a role in shaping

[199] For an in depth discussion see Mele (2003).
[200] Of course probabilistic causation prohibits true control, as we have discussed – so this will not do for us.

the outcome of deliberation, and how we act.[201] All we will do is to sketch a response to this problem; one that I hope is plausible. The details would be complex, and issues of character connect to those of virtue and normativity, as well as practicality. So, we will gloss over some issues, but if we have a plausible avenue for development, then we can at least have some confidence that this problem is not insoluble for non-causalists.

Ekstrom (1993) proposes that one way to think about character is as a web of preferences (which we should recall are a sort of second-order desire). The imagery of a web is helpful, since coherence plays an important role in how agents authorise and assess new preferences and desires. Although we should not adopt her probabilistic event-causal approach, the illustration of a web is one that is suggestive, and helpful. We can conceive of an agent's various motivations, desires, beliefs about the good, evaluative judgements, practical judgements, goals, ends, intentions, plans and reasons, as an interconnected three-dimensional web. The possession of this web is part of what constitutes being an agent, and in particular, being the person they are, with the character they have. When we act, decide, judge and so on, the influence of the structure of this web is involved because this web is partly constitutive of the self that does the deciding. Some items in this web are well connected, and deeply embedded. For example if I have made an evaluative judgement that *X* is of great worth, then this will involve connections with various parts of my conception of the good (which provide a background for even making evaluative judgements), my beliefs about *X*, any desires, intentions and so on, that relate to various exemplifications of *X*, and with any ends or goals whose pursuance would be likely to increase my possession of, exemplification of, or promotion of *X*. The judgement '*X* is of great worth' has a large influence in my character, because it connects so much of my 'self' together, and is deeply embedded in who I am. Other items in the web can be less connected, another evaluative judgment about '*Y* being valuable' might only connect to a few possible ends, and might be something that I find connects to only one of my desires. Now consider my failure to act, or judge, in a way that promotes *X*. More would be at stake for me than a failure to act in a way that promotes *Y*. More of my desires would be frustrated, and perhaps more of my plans would need revising, and so on. Moving or changing items in the web can be more of less 'difficult' depending on how deeply embedded they are.

[201] Ekstrom (2000) poses this challenge, and it is also noted by O'Connor (2000). Ekstrom responds with probabilistic causation in her event-causal libertarianism, O'Connor makes a speculative proposal about carried propensities for certain actions, but his struggle is to explain why this is not a form of non-agent-causal control. In many ways my proposal is a development of insights from both. See also Ekstrom (1993) and (2005).

However, this interconnectedness is not the only important structural feature of this picture. When agents evaluate an end, or involve a desire in deliberation they do not exhaust their consideration of how potential choices can affect them. We are simply not able to calculate, deliberate, or evaluate, in a way that exhausts all of our interests, all of what could be at stake. Even if we are capable of weighing every relevant factor, it would simply be impractical to suppose that we do so on a regular basis. Some concerns are more prominent in our considerations; they have more of our attention. In order to picture this imagine that the web, laid flat on the table, now has some items pulled upwards. So the web starts to look like a three dimensional landscape with mountains and valleys. Something like our desire not to feel thirsty, or hungry, might be a plausible candidate for items we would find at one of the highest points. These high points do not represent things we care about most, that is a matter of being strongly connected to items reflecting my conception of the good, desirable and valuable, but represent items that are prominent in my thinking, ones that I do not need to concentrate to recall, but spring readily to mind and feature regularly in my interactions with this web as I think, deliberate and so on. They are items that are uppermost in my mind. My goal of remembering what I was going to add as a further footnote two chapters ago is plausibly lost at the bottom of some valley.

Now as a matter of fact, when we reflect on the things that are uppermost in our mind, we would expect to find deeply embedded items featuring on these mountaintops. After all, if something is well connected, then it will feature more regularly in my thinking, even if not directly, and by dint of regular interaction is prominent in my psychology. Note also, that as I am conceiving matters, when one item is raised in its altitude it drags the rest of the web with it, and so closely connected items are also elevated in our mind, and some items may be elevated by their connection to several disparate and otherwise non-connected items of prominence. I am conscious that I have used terms like 'mind', 'psychology', 'thinking', and so on rather interchangeably. I do not intend them to be understood in anything like a technical sense for the purposes of this sketch. I only mean to gesture to whatever 'theatre' it is within which the elements of who we are interact with the world around us, and the possibilities this entails. Note that although I have used terms like 'deliberation', and think, I am not committed to the view that consciousness is a necessary tool for interaction. Although the role of consciousness is important, there may be much about the way that my character influences me, over which I am not reflectively aware, such as implicit biases.[202] So when we say that something is prominent in my attention, these need not be conscious attention, it could be non-conscious or sub-conscious attention.

[202] See Levy (2012) and (2014).

However in paradigm cases we are normally interested in instances of thought, reasoning, deliberation and so on, which involve consciousness.

Now, consider a case of and agent deliberating in order to make a decision about how to act. What I want to suggest is that not only does the structure of the web, the agent's character, influence how they decide, but that there is a reciprocal influence. When we deliberate we do not only recognise the weights of reasons but we also assign weights to reasons.[203] The details of the scope of this influence require some careful thought, since some forms of influence over certain factors would be rather implausible – for example can I influence my beliefs by what I do?[204] However there is one aspect of this reciprocal influence that is more easily grasped. When I act for a reason, that reason, and closely connected items in the web, receive a little tug upwards. They become more prominent in my thinking and interaction with the world. Partly this is because we rarely act in isolation from larger plans, and so when we act we commit ourselves in other ways, which may increase a reasons' connectedness in the web. But it is also a practical fact that as a result of deciding a certain way, acting a certain way, and so on, those features of the web connected to the reasons we acted for are ones that are more likely to feature prominently in future interactions. This should not be an alien notion; after all we are familiar with developing habits of thought and behaviour.

The other aspect of this reciprocal influence that is worth noting is that of assigning value to ends (and other subjects of evaluative judgement). If we have a somewhat developed conception of the good, then our evaluation of how valuable an end is will depend on how well that end coheres with the things we have already assigned value to. However our formation of such a judgement is not determinate, nor is it exact. Part of our power as agents consists in not only settling how we act, but also what value we place on things, and what judgements we form.[205] This is not only in cases where we reach an evaluative impasse between incommensurables (as is sometimes suggested), we can do so with impunity. However when we do alter the value we place on something, when we elevate its prominence, this involves our not only producing an effect regarding this evaluation, but also the necessary restructuring that would allow this evaluation to cohere with the surrounding topography of connected factors. There is a cost, and the greater the change in value, the greater the cost.[206] This is a cost of time, mental resources and so on required to effect such a change, and

[203] Nozick (1981) considers this possibility, though in the context of a quite different theory of action.
[204] I leave this a genuinely open question, as there may well be some sorts of beliefs that can be changed by what we do.
[205] I will not re-describe how this is an instance of control, our early comments are equally relevant to these exercises of agency.
[206] This suggestion is found in Bratman (2000).

given the complexity of connections in our web, will often be a significant cost. So we have power over our own character formation. This is another aspect of self-determination: we do not only want the determination to be done by the self, we also want the power to determine the self. Exercising our agency in this way is a doing, and just as practical considerations limit how much we can do when we move our bodies, they limit how able we are to change ourselves.[207] This helps us to understand why agents can have the power to reform their character, but also why their character is somewhat stable, and does not radically fluctuate. On a more fine grained scale, it helps us understand why our beliefs, desires, valuings, and so on do not fluctuate wildly, and why those that are more central to who we are more resistant to reformation. For instance it helps us understand why an addict finds it more difficult to change the way they 'think' about drug use, in fact often very difficult, than a new casual drug taker. It is not that the addict cannot 'think' or do otherwise.[208] This also makes sense of why it is easier for us to change ourselves in our formative years. In our early development, our web will be considerably less complex. When we reach young adulthood, we start to engage the world, and ourselves, in a more self-reflective way. Our experience becomes wider, we are less sheltered from the effects of our own actions, and we begin to utilise long term, intentional plans more as we need to commit ourselves to people, careers and such like. So our web quickly becomes vastly more complex.[209] It is not that we cannot reform ourselves later, but often such major reform will involve our discovery of the good in something that we did not know about before, perhaps, for example, in moral or religious conversion.

Understanding why it may be more or less difficult for us to change an evaluative judgement helps us to see why it is also more difficult for us to act contrary to our most prominent reasons. When we do act against prominent reasons there is a cost, as resources are utilised to effect self-change. However, this does not, or need not, do all the explanatory work. It is also important to note that intrinsic to a goal is something we value that goal for, and thus, something we desire to be satisfied by achieving that goal. It is part of our nature that we enjoy the satisfaction of our desires, but the more prominent our desires, the greater the satisfaction. We do not need to say that the stronger reasons exert more causal influence. A similar explanatory result is achieved by just recognising that we generally develop into persons who seek satisfaction.[210] If we generally acted in

[207] There is some psychological evidence that mental doings do involve the use of limited resources, see Muraven & Baumeister (2000).
[208] See Pickard (2013) for a discussion of philosophers' frequent mistreatment of addict cases, and the power of agents to change their mind.
[209] Perhaps Aristotle had something like this in mind when he argued that virtues, and the Nicomachean Ethics, are for men and not boys?
[210] Or beings, since this point is also relevant to animal agency.

ways that frustrated our strongest desires, we would derive less satisfaction from acting on our reasons. In fact, if we generally acted in a way that frustrated our stronger desires, or at least disregarded their practical importance, then we would undermine our capacity to find ends desirable. It is built in to our capacity for practical rationality that we care about how we act, and that we care about which ends we align ourselves with, our conception of the good, and so on. This 'care' is not causal, it is practical, and plausibly it is something we develop as we seek to interact with the world around us effectively.[211] We develop a standing aim to be effective, and successful, in our interactions with the world. Although this is only a brief sketch, hopefully this gives us some indication as to why possession of this standing disposition, or tendency (avoiding causal reductions of these terms), along with understanding what is involved in acting for a reason, explains why we so often act on our stronger reasons, without needing to give reasons a causal role in action production. So, the non-causal analysis of freedom of choice appears to be coherent, to have explanatory power, and to be defensible. This gives us good reason to take the conception of freedom we have developed as not only capturing, or elucidating, our intuitions, but also as being a conceptions of freedom that can and does obtain for agents. Now we can apply these conditions of freedom to our theological puzzle by considering what sorts of divine interactions agents could sustain which would secure their sinlessness without breaking these conditions.

[211] it is worth noting that the structure of how we value, and what we value, may be a useful as a tool for explaining what it is to be virtuous, or to hold a certain virtue. On some accounts of virtues our caring about practical success is enough, though for a more 'morally motivated' account we might need to introduce a normative concern to be successful at pursuing things that are actually good.

5. Heavenly Sinlessness and Significant Freedom

So, we have discussed what it takes for an agent to act, and for those actions to be free, and we have defended this conception of freedom's coherence. Now we have the pieces of the jigsaw required to answer the questions we developed in chapter one, and in this last chapter our task is to assemble them, and apply our analysis to the theological puzzle of God's preventing sin in heaven.

What does it take for an agent to act freely?

Let us remind ourselves of the criteria we have developed for free actions. We do not need to explain all of the details again, so this is only a brief summary. We were originally interested in what it takes for creatures to possess aseity, which we later related to freedom of choice. But an agent's freedom of action required them to possess more than just freedom of choice, not least because there are also conditions that need to be met for an agent to be performing an action. First of all we characterised actions as an agent's active 'doing', and gave the following conditions,

IV. An agent must possess a desire (or other motivation contributing state), for a particular end, E.

V. We must be able to describe the agent's action in terms that identify it as aiming at this same end, E.

VI. The agent must be aware of some feature of the world that provides a reason for them to take their action as a means to possibly achieve this end, E.

Agents who act, perform something that is goal-directed, where they are motivated to achieve the end which they direct their action at, and are relevantly connected to the world, so that they have a reason for their action that is grounded in the way they, and the world, actually are. We noted that agents can possess freedom of possibilities. But this has to do with the range of possible actions an agent can *successfully* perform. Since the elimination of sin requires agents to not even attempt sinful actions, God's preventing saints from succeeding in attempting a sinful act will not help us explain how to prevent sin in heaven. However we do want agents to possess significant freedom, which means that agents need to have access to significant possibilities. So once we have looked at how to prevent sin, we will need to consider whether heavenly agents are left with significant possibilities for their exercise of free action.

When an agent is free they should be able to act in a way that accords with their all things considered preferences, given their character, desires, reasons and

so on. This sense of freedom has to do with the power of self-determination, or an agent being the source of how they act, so that the action is expressive of the agent. We termed this freedom of action guidance. If an agent possesses freedom of action guidance then their action must be produced by desires, reasons, and so on, of their own. This condition would be most clearly broken if an agent were somehow made to act in a way that was against their preferences. If this were the case then the agent would not be free to act according to their judgements, and it would not be the way the agent is rationally and motivationally constituted that governs what happens. So,

VII. The actions of free agents must accord with the way they are rationally and motivationally constituted.

But we also considered the way that agents end up constituted as they are, particularly if their motivations and reasons are formed under freedom-reducing duress. It is possible to interact with agents and influence their motivations and reasons; in fact, it would be difficult to avoid doing so to some extent. In this regard, it is when agents are influenced in a way that manipulates their motivations and reasons in ways the agent would judge as bad that we are concerned that there is a loss of freedom. Agents need to be able to form reasons and motivations that they value as good. An agent's valuing something as good is a matter of structuring their conception of the world around them and their relationship to it, and this in turn gives structure to their practical and ethical rationality. Freedom of value-alignment is the freedom of an agent to align their motivational bases, purposes, ends, and the structure of their practical rationality with their conception of what is good. We noted that freedom of value-alignment admits of degrees, since influence admits of degrees, as does the value agents may place on ends. So,

VIII. Free agents must not be so influenced that they are unable to form motivations and reasons for ends they value as good and must not be forced to form motivations and reasons for ends they value as bad.

We then went on to discuss the need for free agents to possess an up-to-us-ness, an active power of self-creation, concerning the way that they act. Agents need to be more than instrumentally involved in their self-formation and choices. Being up to the agent to make a difference about whether something occurs or not we called freedom of choice. Our discussion of freedom of choice drew on themes from our discussion of aseity in chapter one. It involves agents in the active exercise of a multi-way power over how they act. It is the exercise of this power that secures the agent autonomy from past determining influences and means they can be active in securing how they act, and not passive. It marks a free agent's ability to control their acting one way rather than another. The exercise of this power is incompatible with determinism and also requires the existence of

alternative possibilities concerning possible actions the agent can perform (which can include refraining from acting as an alternative). But we also argued that freedom of choice can be derivative. An agent's action can be said to be one over which there was freedom of choice, and hence control, not only if there were alternative possibilities that the agent settled at the time of their acting by their exercise of this active power, but also if the settling of alternative possibilities occurred by some past exercise of the agent's active power in acting. This highlighted the need for care when using the presence of alternative possibilities to discern whether agents are free. There may be a great many things agents cannot currently do, or have no reason to do, which is not a limitation on their freedom, but a result of their exercise of freedom. Simply stipulating that free agents need to be able to do otherwise (remember our example was Joe's doing otherwise than agreeing to move in with his girlfriend) is not a good way of articulating what it is to be free. So,

IX. Agents possess freedom of choice regarding an action if the action they perform is up-to-them through the exercise of an active power to settle among alternative possible action, concurrently, or in the past.

We should not need to say much to remind ourselves about an action constituting an agent's application of their capacity for means-end rationality in their behaviour, since we covered this in the last chapter. But we need to highlight again that agents can possess a multitude of reasons and desires, and they exercise their freedom of choice over how, and when, they act on those reasons, or satisfy those desires, by acting – which is intrinsically doing something for a reason. Importantly, agents can act on any of their reasons, and do not only 'act', when they act on their strongest reasons (or motivations/desires). We described agents as possessing (and being partly constituted by) a web of beliefs, desires, judgements, and so on, whose topography models how some features of their character are prominent, and how changes to their character can require a 'cost' in reforming that topography. When agents act against prominent reasons there is a cost, as resources are utilised to effect self-change. When agents deliberate, decide, choose, and act, they do not only recognise the structure of their web, or the weights of reasons, or value of ends, but there is a reciprocal effect, and deciding, choosing and acting alter the structure of the web, or change the weightings assigned to reasons and ends. So the power agents possess is not only one over what actions they perform, but is also a power over the development of their 'self'. Agents possess some power over their character. The illustration of the web with a well-structured topography will be a useful illustration in our discussion.

We can summarise these conditions for free action as,[212]

I. Motivation possession
II. Goal-direction
III. Reason awareness
IV. Character accord
V. Value alignment
VI. Aseity

How can God prevent sin in heaven – the beatific vision

So, God needs to ensure that it is impossible for agents to sin in heaven, while preserving the conditions required for agents to remain free. One of the most prominent themes in theological discussions of heaven is the beatific vision enjoyed by glorified saints. The saints in heaven are glorified, that is to say they experience being brought into the presence of the glory of God.[213] But they are also changed so that they are fit for heaven, and this change is the completion of their redemption. There is a physical change, but more interesting for our purposes, the saints in heaven are perfected in their nature and character, so that they are without blemish or blame.[214] When the redeemed reach this state, they have been glorified.[215] Roman Catholic Christianity also has a doctrine of purgatory, which is a place in which God perfects the redeemed prior to entering heaven.[216] However this will not interest us, as we are interested in what the change to a glorified state consists in, rather than when it occurs. So, purgatory or not, by the point of their glorification, God has effected a change in the redeemed. In the presence of God's glory the saints experience the beatific vision, which is a particular experience of seeing, and knowing God. Since this is a key idea, we will

[212] These conditions, and the statements summarising them earlier in this chapter, are not supposed to be taken as rigorous definitions of the criteria, and are only devices for ease of reference. For the details, we must refer back to chapters two to four. Note also that I have avoided some well-known terminology, in favour of these terms, as I do not want to import additional theoretical baggage, for instance IV is close to Timpe's (2008) 'sourcehood', or Fischer & Ravizza's (1998) 'reasons-responsiveness'.

[213] See Romans 8:30. Also, note that 'the redeemed' refers to all humans whose destination is heaven, and 'the saints' refers to the human occupants of heaven, not some special subset of the redeemed. For a brief discussion of glorification see Demarest (2006), and Elwell (2001).

[214] See Colossians 1:22, Ephesians 1:4, Jude 24, and 1 Corinthians 1:8 & 13:10, where the focus is moral perfection.

[215] Eastern Orthodoxy has a similar concept, but call it deification, but we will not discuss the theological differences here.

[216] See Kreeft (2001:149 ff.) and Walls (2002), though the later is focused more on issues surrounding salvation.

consider its basis in Scripture to help us clarify matters.[217] There are three key texts,

> 1 Corinthians 13:9-12, 'For we know in part and we prophesy in part, but when the perfect comes, the partial will pass away. When I was a child, I spoke like a child, I thought like a child, I reasoned like a child. When I became a man, I gave up childish ways. For now we see in a mirror dimly, but then face to face. Now I know in part; then I shall know fully, even as I have been fully known' (ESV).

> 1 John 3:2, 'Beloved, we are God's children now, and what we will be has not yet appeared; but we know that when he appears we shall be like him, because we shall see him as he is' (ESV).

> Psalm 17:15, 'As for me, I shall behold your face in righteousness; when I awake, I shall be satisfied with your likeness' (ESV).

The beatific vision provides the saints with a knowledge of God that is different to the knowledge available on Earth. The translation of *ἐπιγινώσκω* as 'fully' is a popular, but unhelpful translation, as it seems to indicate that we will know everything about God. However, God is infinite, and we are finite, and thus limited in our cognitive abilities. So it is impossible for us to know God fully in the sense of comprehensively. The force of the preposition (*ἐπι*) can also be rendered as 'exactly'.[218] This contrasts with the idea that our earthly knowledge of God is limited by some imprecision, or uncertainty, as the context 'seeing in a mirror dimly' suggests, and is also borne out by the other texts. Giesler (2004) describes the difference as our earthly knowledge of God being indirect, and our heavenly knowledge being direct. On earth we are at an epistemic distance from God, and there can be epistemic uncertainty. This means we are, to an extent, justifiably uncertain about some of our knowledge of God. However, in heaven, there will be no uncertainty, and we will be fully convinced of what we know. This may enable heavenly beliefs about God to play a greater role in our thinking. Such beliefs would have the same sort of status as undeniable truths, ones that we cannot deny, contradict or question (as long as our capacities for thinking are in good order). There is also an indication that we will know more about God in heaven than we do on Earth, such as the contrast between our 'knowing in part' being a state that will pass away. So there may be things that we do not, or even cannot know about God, but will have access to in heaven.

So, in heaven we will have more true beliefs about God, and experience of knowing God directly, and will have no uncertainty concerning our beliefs about God, or our experience of God. We do not need to explore what the content of

[217] Generally, we will try to avoid as many questions of exposition and Biblical theology as we can.
[218] See Danker (2000) – BDAG 3rd ed.

these extra beliefs could be at the moment. But we can ask whether this beatific vision, and especially the knowledge of God it imparts, could be something that would prevent sin. First of all, it is sometimes suggested that it is impossible to sin in God's presence because there is supposed to be something overwhelming about God's presence, which so captivates creatures that they become unable to do anything other than obey. We would still like to know why this occurs, as this is not in itself an explanation. But it turns out not to be a helpful approach anyway, since it is in fact possible to sin in God's presence. The Devil, and angels, were created in God's presence, and while there, sinned and fell. If these free creatures were not overwhelmed into obedience, then neither would the saints be, unless there is some other difference between the two. So being in God's presence is not sufficient to prevent sin. However this does not mean that there might not be something important about being in God's presence. Perhaps the presence of God is required for it to be possible to not sin, so it could be a necessary condition.[219] Even if it is not a necessary condition, and it is possible to not sin on Earth, it certainly would seem to be a relevant motivating factor – we are all aware of a general tendency to do the right thing when we are being watched.[220] However, it is not a tendency that always works, and we need it to be impossible to sin.

Could the knowledge the saints possess somehow alter their range of possible actions to prevent there being any sinful possibilities? This approach asks whether there are just no sinful alternative possibilities given the way the saints are constituted – their reasons, beliefs, desires and so on. An alternative approach, which we will come to in a moment, is whether the constitution of agents does leave sinful alternative possibilities, but for other reasons they are possibilities that cannot be acted on (so they are conditional possibilities whose conditions are never met). The knowledge an agent possesses will provide the agent with beliefs about ends, some of which may be ends that the agent then desires. What the agent knows will also give rise to reasons about the means to achieve certain ends. So if knowledge is going to prevent sin, it must do so by providing the saints with a rational and motivational base that cannot produce any sinful act. Consideration of the first free creatures is again illuminating. The Devil, and Adam, did not possess any false beliefs, or bad desires, when they were created. All that God made was good, or in the case of Adam, very good. This much is stated in the creation text, and is also a theological expectation – a perfect being should create the best possible creatures. So it seems plausible to say that Adam and the Devil did not possess anything in their rational and motivational base that shouldn't

[219] We will not explore this, but there may be ramifications for original sin here, as well as the question of why there is no redemption available for angels who fall.
[220] Although this earthly analogy doesn't quite capture what it means to be in God's presence.

have been there, for example a desire for something that it would be wrong to desire, or a reason to act in a way that would be wrong. If they did possess some flaw in their character, then they might rightly complain that they could have been better made.[221] That is to say, something could have been done to prevent them sinning, and this would jeopardise the free will defence. So our commitment to the free will defence, and our theological expectation mean that we have in Adam and the Devil, free agents without any flaw in their rational and motivational base. And yet, they fell. So if it is possible for free agents with unflawed characters to fall, then any knowledge gained by the saints, via the beatific vision, does not look like it will be able to prevent the same possibility.[222]

This does raise the question of why Adam and the Devil did fall. We do not have space to fully explore this issue here, but I will say something, even if a little speculative, since it will help us to think about the agency of Adam, the Devil and the saints.[223] Anselm's *De casu diaboli* is, arguably, the most important work on this question. Anselm claims that humankind lost rectitude, the desire and ability to rightly order desires, at the Fall.[224] Rectitude is then restored at glorification. Since rectitude is supposed to preserve the saints in their sinlessness, one might wonder why it was ineffective prior to the Fall, which means that rectitude alone is not something that will do all of the explanatory work.[225] However the idea that free creatures need to possess more than just good desires (and reasons, etc.) in order to do what is right is interesting. Creatures need to apply their desires in the right way and can fail to do what is right if they fail to apportion each desire, or reason, its correct place in their deliberation. An example may clarify. In the Fall account in Genesis Eve takes the fruit after being tempted by the serpent to do so. But the serpent did not need to appeal to a sinful desire to take something that was

[221] I am using 'character' as shorthand for a rational and motivational base.
[222] The Scholastics wrestled with how to explain the role of knowledge and love in securing sinlessness (understandable, given the prominence of the natural law tradition). We do not have space to exposit and examine their views, but see Gaine (2003) for a discussion of their approach, especially, Aquinas, Suárez and Scotus.
[223] Timpe (2014) contrasts intellectualist and voluntarist accounts of the primal sin in scholasticism, concluding that neither has the upper hand on the question of arbitrariness, which is, in a way, just the luck objection in a different context (in fact, our non-causal account can be seen as an attempt to synthesize some of the intellectualist and voluntarist concerns, but relating things to their concerns must be left for another time).
[224] It is difficult to relate Anslem's discussion to modern categories of analysis, and it is unclear whether Anselm saw rectitude as a power that was distinct in its own right, or a desire that was an input to some other power, like the power to act, or decide. I suspect the latter, but do not want to get drawn into an expository fight, it is not required for our purposes here to be this accurate in expositing Anselm. For a detailed discussion, see Rogers (2008).
[225] There is a suggestion by Rogers (2008:89) that Anselm claims that ignorance is required for free choice, but then we are left wondering whether this would be 'very good', and in any case, this would mean that there is less scope for freedom in heaven. Anselm's conception of freedom is different from our account, so these differences are unsurprising.

forbidden, nor did he need to introduce a sinful reason or desire to Eve, which she subsequently accepted. In fact, the telling part of the temptation was that eating the fruit would make Eve more like God. Now considering the emphasis in Christianity on human perfection and growth being a matter of becoming more like God, Eve's desire to be like God does not seem to be a desire that is out of place. However, Eve acts on this desire, or for this reason, without balancing her consideration of what was appropriate to do with other reasons, such as to obey God, or to trust what God has said. The serpent played a part in stoking some uncertainty in Eve concerning reasons that could have been brought to bear on her deliberation. But this is illuminating for our problem. Eve possessed the ingredients for right action, but these ingredients did not issue in the right action. We noted that as free human agent we possess cognitive limitations concerning our ability to include all reasons in our deliberations, and we also noted that we possess the ability to act on any of our reasons (often more than one). This ability is required for us to possess aseity. So perhaps Eve's cognitive limitations create the space for uncertainty to be possible about how her reasons apply to the alternative possibilities before her. Furthermore, since she possesses aseity, she can adopt any of her reasons in acting, and her acting for the reason of 'wanting to become like God' satisfies all of the criteria for being a free action. So Eve's sin was deliberate and not an accident.[226] She was not forced to sin and possessed the ingredients required to act in a non-sinful way. But due to the nature of freedom, she did not possess any further ability to order her reasons in a certain way. A similar story would need to be told about Adam and the serpent, but we will not attempt that now.

There is certainly a lot more to be said about the psychology of temptation, and Biblical scholars may complain about the exegesis of Genesis here. But it does at least illustrate a point about what free agents can secure and what they cannot, and that is important for us.[227] All of the ingredients for sinlessness can be in place in an agent's psychology, in their rational and motivational base, and this makes performing a non-sinful action possible – but it does not make performing a sinful action impossible. Clearly agents can sin if they possess a reason to sin that is itself sinful, such as my desiring to kill someone because I hate them, and my holding this knife behind them providing a means to satisfy this desire. But agents do not need a reason to sin in this sense because sin can also be a matter of misappropriating their desires and reasons, which are perfectly good in the right contexts. So if we are going to secure sinlessness, we need to do more than remove flaws in the characters of agents.

---

[226] Contra the objections that it was arbitrary, discussed in Timpe (2014).

[227] A helpful starting point for how temptation engages an agent's practical rationality is Bratman (2008).

But perhaps there is some form of knowledge that is only available to those who have fallen and been redeemed. This would mark a difference between the saints and Adam, and so could explain why what they know keeps them from sin, while what Adam knew could not. There is a plausible candidate for what the saints know that Adam did not (in his innocence). Adam had never sinned, and so he did not know what it was to feel guilt, shame, pain, separation and other experiences contingent on sinning. Similarly Adam did not know the joy of being forgiven, the humility and repentance required for reconciliation, what it is to be the recipient of mercy, and so on. Before we even consider whether this knowledge could prevent sin, there is a question about whether God could have created free beings with this knowledge. Presumably God knows what guilt and shame are, as well as forgiveness and repentance, even though God has never done anything to be ashamed or guilty of, or that needed forgiving or repenting from. So, should it not, in principle, be possible to provide creatures with this knowledge without their having to experience sinning? But perhaps it is not the propositional knowledge 'that guilt feels bad in this way', or 'that being forgiven brings joy in this way' that is the difference we are looking for. The redeemed relate to the content of these propositions in a special, first-personal way. Not only are they aware that sin is bad, or that guilt and shame have certain properties, but also they are aware that they were sinners, that they felt guilt and shame, and were transformed by experiencing forgiveness and reconciliation. The difference is analogous to the difference between knowing that being in debt to someone would carry certain obligations, and knowing that I am indebted to you, and have certain obligations. It shifts the focus of analysis from the properties of certain states, to an analysis in terms of relationship (one they are in), and the effect that certain events have had, or could have, on that relationship. However, an objector might retort that God could perhaps have provided such knowledge, or awareness or what being in these states is like, through some other means, such as giving agents a vision or dream. But let us assume, for the sake of argument, that there really is something that the redeemed have that Adam not only did not have, but could not have had. Is this extra knowledge sufficient for preventing sin?

The redeemed know that they did sin and were forgiven. The have experienced the pain, guilt and shame of sin. It is very plausible to suggest that such knowledge furnishes an agent with an extra reason not to sin. After all, burning myself taking a pot out of the oven gives me a reason not to remove pots from the oven without a towel in future, and the burn, or memory of it, or the sight of the scar, all provide me with a motivation to avoid that action. Perhaps the memory of having sinned, and then having seen sin for what it is, the horror of the pain it causes God, the loss of potential goods that then cannot be realised in our relationships, all seen with a clarity that the beatific vision imparts, is like a

vivid scar on the character of the saints.[228] The possession of psychological states like this would certainly be helpful for preventing sin for the saints. But does knowledge like this necessitate my never sinning again? The problem, analogously, is that no matter how bad the scar, I still do sometimes burn myself on pots from the oven, or rather, can still so burn myself. Given our discussion above of the capacity to misappropriate our reasons, this is not surprising. So this extra non-Adamic knowledge could help prevent me from sinning but is not sufficient for doing so. However, there is something positive to draw from this proposal. If God could somehow ensure that I was always reminded of this scar and that it always effectively influenced my deliberations, then it could be something that God could use to ensure I don't sin. This might be a good thing, because if the only other alternative was to never let me near an oven again, or to always be guiding my movements when near an oven (like holding the hand of a child), then I might miss out on the opportunity to enjoy interacting with an oven in a good way. The analogy is becoming a little stretched, but we will pick up this idea momentarily.

The idea that knowledge imparted by the beatific vision is not sufficient to prevent sin could create a problem for us when we consider the issue of divine hiddenness. If the beatific vision overwhelms creatures, so that they are constrained to love and obey God, then this supplies a reason why God must provide creatures with an epistemic distance in order for them to have opportunity to freely choose to love and obey God. Our discussion might lead us to note that this is not a wholly tidy arrangement of ideas, for Adam walked with God in the garden and may not have experienced divine hiddenness in the same way human agents do after the Fall. But putting that to one side, the possible objection is that our claim that the beatific vision still leaves agents free to sin leaves us without a reason for divine hiddenness. It is implied that we would really like a reason for divine hiddenness; because God's apparent epistemic scarcity troubles us. But we can propose an alternative reason. Having sinned, if God were to be present in the way he is in heaven, and we were to see ourselves, God and our sin for what they are, we would be unable to countenance the idea of approaching God and asking for forgiveness. Put another way, the beatific vision would see us consumed by guilt and an awareness of our shame and separation in relationship with God. This would make it difficult for us to approach God, or consider turning to God.

We have two options here. On the one hand we could claim that such an overwhelming sense of separation would make our turning to God (and thus being redeemed) impossible, but that we would still be free as this would only limit our

[228] The redeemed are healed, but not cured.

doing one kind of thing, and we would still possess the capacity to be free regarding other possibilities. This might be a difficult position to take, but the resources we develop in our answer to how sin is prevented may help to defend such a position. But if it turns out, for some reason, to be a position whose costs outweigh its benefits, we can still defend divine hiddenness with a milder claim. Perhaps the sense of separation does not make turning to God impossible, but it makes it very difficult. This is analogous to the issue of our knowledge of the scar of sin making it more difficult to sin in heaven, but still a possibility. Now, given that God desires to see as many creatures redeemed as possible, and the salvation of more creatures is a good thing, if there is anything God could do to make turning to God less difficult, then he would have a moral justification for so doing.[229] This could explain why God is distant between the Fall and heaven. It provides a little breathing space for creatures to consider and explore their lives, God, sin, and so on, without being immobilised by an awareness of their sin and it's effects.[230] There is more to be said about the balancing act required here, as agents would need some awareness of sin. Understanding how distant God should be, or whether there is space for closing the epistemic gap on some occasions (such as visiting angels, and visions) raises some difficult questions. But I think we have said enough to see that this issue will not be fatal to the position we are developing.[231]

How can God prevent sin in heaven – union

So, we need something in addition to the beatific vision to secure that saints necessarily do not sin. One avenue to consider would be Christology. Jesus possessed a full, free humanity, and yet did not sin, because he was necessarily impeccable. So if we understood why Jesus did not sin, that might help us understand the sinlessness of saints. However this may not be as helpful as it might first appear. The person, Jesus, could not sin, because he was divine and thus necessarily sinless (we will not try to explain why God cannot sin). The humanity of Jesus' dual natures meant that Jesus possessed the capacities relevant to the possibility of sin when they are considered apart from the influence of the divine person (and nature). The influence of the divinity on the humanity is labelled *perichoresis*. This interaction made it possible for there to be a human nature that was necessarily sinless. However, for our purposes we are interested in

[229] Such help, Suárez calls congruent grace, see Craig (1989).
[230] Our discussion has been framed in terms of salvation and turning to God, but this approach could be extended, and there may be other aspects of flourishing in our relationships with God, and other creatures, that benefit from (or, on the previous option, require) some divine hiddenness for a time.
[231] For an excellent collection on divine hiddenness see Howard-Snyder & Moser (2002).

exactly what this influence that the divine nature exerted on the humanity consisted of. So, understanding the effect of *perichoresis* when it comes to preventing sin, is not a solution to our problem, but the reoccurrence of our problem in a different context. Christology may be helpful for understanding the means of this influence, but that is not central to our inquiry. We could just label the means of God's interaction with the saints as being by supernatural means.[232] Even in a Christological setting, we would still need to examine the capacities essential to free human nature, and how they can be interacted with to prevent the possibility of sin.[233]

So, in order to prevent sin, there needs to be something extra, added to the saint's character. We saw that glorified agents may include some new prominent features in the topography that represents their character, such as their hating sin, recognising the pain of sinning, and so on, as well as positive reasons/desires such as the joy of forgiveness, their awareness of the need for dependence on God, and so on. Let us just sum up whatever these new features might be as a 'love for God', which theological tradition suggests stands a good chance of being at the root of all of them anyway. These features might be new in the sense that they are a fresh revelation of something never before apprehended by the redeemed, or could be new in the sense that they are an amplification of the awareness of some issue. The problem is that a prominent love for God does not necessitate never sinning. So the agency of the saints needs to be augmented in some way, to ensure that the love of God remains prominent and always has an appropriate effect, so that it is never a reason whose application is missed or diminished when agents act. This means that this something extra must involve an interaction that limits how agents act for reasons, which means that God exercises some control over how agents act. This influence has two aspects, but they are two sides of the same coin. Recall that we said that an agent influences their character when they act, and different reasons, ends and desires can become more or less prominent when acted on. So agents need to be prevented from acting in a way that would mean that the love of God was made less prominent, though other features of who they are could change. As long as the love of God is prominent relative to the agent's other reasons (and dominantly so) then it exerts greatest influence over the way the

[232] For the record, the role of the Holy Spirit in the perichoresis found in the natures of Christ is relevant to understanding the 'means' question, and thus an exploration of this issue would be illuminating for the development of a metaphysics of sanctification, or as the Eastern Orthodox theologians would have it, deification. This would lead us to develop a more Trinitarian picture of the relationship between the saints and God, and the different roles each of the Godhead could play in securing sinlessness. But, though interesting, this is not an issue we have space to explore. See Crisp (2007) & (2009), and Sanders & Issler (2007).

[233] I have tried to say only a little about this, since Christology raises so many metaphysical puzzles, and have only gestured to an avenue for future research. A good introduction is O'Collins (2009), and Rae (2009).

agent orients themself towards the world (of heaven) and the way they order their thinking, deliberating and evaluating of it. The structure of their web is the lens through which they interact with the world. However, this structure does not only present weights of various reasons, values and so on to the agent, as the agent also has the power to alter the structure of their web, and they do this by exercising their agency in acting, choosing, forming intentions, making plans, and so on. We noted that there is a cost in such restructuring, and so some restructuring is more difficult than others. So agents also need to be prevented from applying their reasons, values, and so on, to the exercise of their agency in a way that does not feature the love of God as a dominant reason for what they do.[234]

At this point, matters become very speculative, as we are not given many clues about how God might effect this influence. However, the theme of permanent unity between God and the saint is a prominent theme in Scripture, especially in the sense of God becoming a part of human psychology (often using the term 'heart').[235] This is expected, since we have already argued that God cannot provide an external prevention of sin. When agents do act in a way that requires a change in their self we noted that there is a 'cost', or we could think of it being difficult, or requiring more effort, to act against a habit, or central character trait. If God is acting upon the web of the saint (their structured set of beliefs, reasons, etc.), God can ensure that this 'cost' is always above a threshold that means that sinful applications of the agent's rational and motivational basis through their acting are off limits to the agent, or beyond their ability to produce. In the same way that an agent has the power to change their 'self' and promote or demote ends, reasons, desires, and so on, God can also exercise his agency to promote and demote facets of the agent. God could achieve this by the sort of normal influence of interacting with the agent, for example by speaking to them. In the same way, I could influence a friend to do something important by reminding them how important it is to them. God's acting in this way can be guided by God's perfect knowledge of the agent, that is to say, God would always know what to say and when. But the divine influence must go beyond this, for God needs to ensure that agents always listen and apply what they are told in the right way. So there must be a form of divine influence over agents that exerts a promotion of some features of the saint's self, or blocks what would have been the promotion of something that would have then been out of place, by exerting an influence counter to that of the saint. Thus the power of change over the self is

[234] In other words, whatever heavenly agents do, it should be expressive of their love for God, perhaps not explicitly so, but since their web is interconnected, however they act will be oriented implicitly at the end of loving God.
[235] See Jeremiah 31:33, Ezekiel 18:31, Romans 5:5, 2 Corinthians 1:22, Galatians 4:6.

one that the agent possesses in cooperation with God.[236] God does not take over, but is a contributor to self-formation.

This second requirement is more metaphysically puzzling, but I raise the issue of the first form of influence to make the point that supernatural divine influence of the second kind might be something that is rarely required. In the first place, the change effected by the beatific vision leaves agents strongly oriented towards acting rightly and evaluating the world as it actually is. So we would expect the times when intervention is required to be rare, though they will occur. But also, much of what agents may need to keep them away from sin is by way of maintaining a healthy character. Again this will require some times of intervention, but much of this maintenance could be by God showing things to the saints, talking to the saints, and the saints talking to each other. This sort of interaction is no more a threat to freedom than my reading a book on how to be a good parent. In fact, it may even be a good thing that much of the saints' maintenance is via explicit interpersonal interaction, since relationships are largely constituted by two-way interpersonal interaction, and these forms of interaction may instantiate something valuable, which makes a relationship good.

But rare though they may be, we do need to ask whether these special, supernatural acts of intervention are bad, or freedom destroying. When an agent acts in a way that is cooperative with God[237], that is to say, includes divine influence, are our conditions for free agency broken? The agent will still possess a motivation for an end, since their action is the application of one of their own desires and they will be acting in order to achieve an end that they themselves possess, and the performance of the action will satisfy them. God does not need to implant an alien desire or end; he only needs to manage the possible ends that agents aim at. This could be by preventing one choice through supressing a desire or reason, or by ensuring the better choice by promoting the relevant desire, or reason. Agents already have the ingredients required for good ends through what they have learned by being glorified. Therefore motivation possession (I), and goal-direction (II), are both satisfied. Reason awareness (III), is also satisfied, since agents, once adopting an end, or when adopting an end, do so because they are aware of their heavenly environment and know (in a way that does not admit of the possibility of error found on earth) that features of heaven, or of God, provide a basis for the appropriateness of their action to its goal. Since the way the agent acts derives from ingredients found in the agent, character accord (IV) is trivially satisfied. But does any of the influence God exerts mean that agents are

[236] I use the term 'cooperative' advisedly, bearing in mind cooperative theories of grace in soteriology, but that is a subject we do not have space to explore, though it is connected.
[237] Or decides, chooses, plans, or any other candidate for a use of agency which may require God's supervision.

unable to form motivations and reasons for ends they value as good, or are forced to form motivations and reasons for ends they value as bad? Well, saints have a perspicuous knowledge of which ends are good or bad, and do not evaluate any good end as bad, or vice versa. God, necessarily, only promotes ends that are good, due to his good nature. So it is impossible in principle for an end God promotes to be one that an agent would evaluate as being bad. This is tantamount to saying that a saint would never judge any of this divine intervention as being unwanted or bad for the agent. Therefore value alignment (V) is satisfied.

The last condition is aseity (VI). Given that some influence over the determination of the agent's action in these rare situations is external to the agent, it does appear that what happens is not up-to-the-agent. It will not do to claim that since the nature of the influence is cooperative, such that the way the agent acts is up-to-'God-and-saint' the condition is satisfied. For aseity was supposed to mark the independence of agents in determining how they act among alternative possibilities. The removal of a possibility by another was a mark of manipulation. But remember that this aseity is not entirely unfettered from all influence each time it is exercised, and the way that agents have exercised their agency in the past can mean that the absence of an alternative can still be something that is up-to-the-agent, even though this is not through the agent's current exercise of their powers. So is the heavenly limitation of an agent's possibilities for the exercise of their active power to choose a limitation that has resulted from a past exercise of their agency? During our earthly existence we have ample opportunity to exercise our capacity in a way that limits future possibilities, but are any of the ways we exercise aseity relevant to this particular limitation, and, also common among all of the redeemed? What about the decision of the redeemed to accept Christ's offer of salvation? This is certainly a necessary requirement on the redeemed, so will be a common exercise of past agency among all the saints. But is it relevant to the limitation of my heavenly possibilities for action? This is a little difficult to answer because it is not quite the same as the case of the drunk driver, which we considered in chapter three. When we considered the drunk driver, his past action (drinking too much) directly prevented him from having the later possibility of braking in time to stop. But my choice to accept salvation does not seem directly linked to my being unable to act in a way that would threaten the maintenance of, and integrity of my sinless character. It requires the intermediary of God acting to get us from the choice to the actual prevention of a possibility. By analogy, this might be like my asking you, who are also drunk, to see if you can do something about the brakes on my car, since I would like to make a bogus insurance claim, so you drain the brake fluid. My later failure to brake is a result of my exercise of agency, but it also includes your exercise of agency, since you did not have to do what I said.

However there is an important disanalogy, asking God to 'cut the brake lines of sin' does not leave the performance of the divine action uncertain, for God has promised to do this, and cannot break his word. So, because of the nature of God, in asking God to save them from sin, the redeemed have necessitated the way that God will intervene, and that God will actually intervene. If we consider what is involved in an agent exercising their agency in accepting salvation, we will be able to see that the sort of heavenly influence the redeemed are subjected to will not be an unwanted surprise. We will only offer some brief thoughts on the nature of this choice, but we can say enough to see why there is plenty of scope for the decision to accept salvation is relevant. In order to come to a point of recognising that the invitation of salvation is something that is worth accepting, the sinner needs to come to a recognition that sin is a problem, and that it is a problem that the sinner cannot deal with on their own. This involves consideration of the need for the forgiveness of past sin and some way of dealing with the prospect of their possibly sinning in the future. The sinner must come to the recognition that their own control is insufficient. The sinner also needs to recognise that salvation is by grace, that is to say, there is nothing they can do to merit salvation. So the sinner acknowledges that God has offered to do something about sin (and the sinner may have only a vague idea what this will consist in). The key though, is that part of what the decision to accept salvation consists in, is to put oneself in submission to God. Having come to the end of one's own efforts to exercise control over oneself in the right way, a sinner relinquishes their autonomy and invites God to participate in the controlling of their life – in their self-formation.[238] So the nature of the decision to accept salvation involves sanctioning God to interact with the exercise of the agent's capacities to act, think, decide, reason, evaluate, and so on.[239] So there is a past exercise of freedom of choice, which is directly pertinent to the limitation of heavenly alternative possibilities, which can be the basis of the claim that the instances of agency we are focussed on, ones where God cooperatively influences the exercise of aseity, are derivatively free. Thus condition VI is satisfied, and God's interaction does not destroy freedom, but is able to prevent sin.

Now, this argument does depend on a particular conception of the role of free agency in salvation. This is a controversial issue, but we should bear in mind that those who are disposed to deny the role of libertarian freedom in salvation are also likely to be uninterested in developing a free-will defence. Compatibilists about salvation are usually compatibilists about evil. So although a defence of libertarian

[238] Although the sinner does not initiate this interaction. God initiates the process of salvation, the sinner's role is to relinquish their resistance, but this is still an exercise of agency. See Stump (2012).

[239] Sennett (1999) suggests that earthly choices provide a pattern of character that enables God to intervene, but he claims that the saints only possess compatibilist freedom.

freedom in salvation is an interesting project, it is not one we need to undertake, since we are already committed to exploring what follows from the free-will defence. A second issue is that there do seem to be some agents in heaven who never had the opportunity to exercise their own capacity of free choice in accepting salvation. There are a number of candidates, but the prominent group are babies who die before developing the capacities required for free choice. It seems harsh, or even evil, to commit these babies to an eternity of separation from God and heaven, and many Christians are committed to the view that God secures the salvation of babies. We should note that this is not the only view though, and it has been the official position of some traditions that God only secures the salvation of the babies of the redeemed, however this does not really solve the problem, since these babies are still saved without recourse to any exercise of their agency.[240] The problem is that, without ever having exercised free choice, whatever God does to these agents in glorifying them, and keeping them from sin, does not look like it can be derivatively free, and then such interaction does seem to be freedom destroying. Responses to this problem face a problem, since saying that the lives of these redeemed babies in heaven are still good (and sometimes it is claimed, free), leads us to wonder why God could not just do this with everyone, and we undercut the free will defence.

Explaining what happens to the babies is, I think, probably the most difficult question in theology. So, unsurprisingly, we will not attempt to solve it. However, since we are interested in the possibility of consistency, if there is a plausible option to explore, this issue will not pose a fatal objection. There are four options, the first two might be the worst-case scenario, but would resolve the problem as far as our discussion goes. Either all babies do remain separated from God for eternity, or God annihilates the souls of all babies (so that they cease to exist).[241] What needs to be supplied is some explanation for why God's doing one of these things is morally justified, and this appears difficult. On the other hand God might supply babies who die with some post-mortem opportunity to develop their capacities for freedom, and to exercise a choice concerning their freedom, securing God the permission to either save them, or not. This possibility lacks any clear Biblical support, but is perfectly consistent. The fourth possibility is a Molinist solution.[242] Perhaps the counterfactuals of creaturely freedom can ground God's interaction with the babies, based on what they would have done. Molinists may

[240] Roman Catholicism suggests that babies may escape punishment, but be denied the beatific vision, but there are a number of theological and Biblical reasons for rejecting this suggestion, however we cannot explore that issue here.

[241] This also raises some questions regarding original sin, but too many philosophical issues will rear their heads for the space we have if we attempt to map how this doctrine interacts with this problem.

[242] There is nearly always a Molinist solution to any theological puzzle!

be able to make this solution work, but will have to explain why God does not do this with everyone or only choose to create those he knows will be saved (or perhaps just accept a form of universalism). But given that a response to the problem of evil, at least one that is libertarian, is going to have to answer the question of why this world, with all the people who are seemingly not saved, is the best possible world God could create, perhaps Molinists will have the resources to accomplish this task, or stand as much chance of succeeding as does the free will defence. I am not going to endorse any one of these options, but only note that there does appear to be hope for a solution, and so the objection is not obviously fatal to our claim about the requirement for the exercise of earthly freedom.

Why could God not prevent sin in the beginning?

Given our discussion of the need for genuine creaturely aseity prior to God being able to interact in a non-freedom destroying, but sin preventing way, the answer to this question is straightforward. God cannot create agents who have decided to submit to God, because an essential property of submission (in our sense at least) is that it is freely chosen. If God created beings who were 'submitted already', then he would have creatures who did allow God to intervene, making sin impossible, and had characters prominently disposed to obeying God, but this would not really be submission. Such beings would have never possessed an alternative possibility regarding their relationship with God and so would not be free concerning it. Such a being would 'do' the things that saints might but would be like one of the radio-controlled agents we discussed when formulating our conditions for agency. In this position such agents would not be submitted but would be dominated. We could express this same idea in terms of other relational terms, for example un-free agents cannot really love God and cannot really worship God. They could go through the motions and appear to be worshipping, but if there is any good or value in the relationship of worshipping, this would be lacking (see the discussion in chapter one).

But would this be such a bad thing? God could do this, and prevent sin, even though he would not have any free creatures. We could suggest that forcing beings to be a certain way would be a bad thing in itself. However, although it is plausible that disregarding the 'rights' of autonomous agents would be a bad thing to do, making robots do things does not seem to be bad for the robots – there do not seem to be any rights. So if sinless created beings were not really agents at all, this would not be an objection. However, we might argue that agency is required for some goods to obtain, and therefore, even though agency regarding sinning might be destroyed, it remains in some other way. In this case agents would be in some way autonomous, and perhaps then in possession of 'rights' regarding any

curtailment, or destruction, of their freedom being a bad thing. However, agents who are forced to be good, in a freedom preventing way, cannot possess a valuable freedom, even if there is a freedom left for them to possess.[243] The reason why this is so, also explains why making non-free beings would be a bad thing, because there would be some value, or good, which it would be good to have, that was, as a result of the way God created, unavailable. The reason is that if freedom is to be valuable, we need to be free to sin and that the value of freedom is a great good.

Consider the second part of this claim first. Whatever reasons God has for creating, given that there is an outstanding ethical question about whether the evil that has resulted from God's creating is justified, they must be reasons that involve there being some good that obtains as a result of creating that makes doing so worthwhile. So is there a reason to think that the greatest goods involve free beings in relationships with other agents? Much could be said about this, but I will suggest only one line of thought to motivate our discussion. God is the greatest possible good. This is, hopefully, uncontroversial given some sort of perfect being theology.[244] So how does God instantiate goodness? Is it by exhibiting astounding structural complexity, or by the vastness of his spatial extent, or temporal duration? No, because these are not properties God can instantiate. So creation cannot be a great good because of the beauty of the galaxies, or its evolution through time. Or, at least, if this is what the goodness of creation consists in, it is far less than the greatness of the great good that God is. Perhaps it is in the instantiation of moral perfections that the goodness resides. Now we have not argued that freedom is required for moral responsibility, but did note that it does depend on freedom in some way. So this would give us some reason to suspect that a greatly-good creation would require free creatures. But the most obvious candidate for the greatness of God lies in the instantiation of perfect, loving relationships between the three persons of the Godhead, relationships that are so perfect that the three persons constitute a unified single being.[245] So a plausible candidate for the great-good-making property of creation would be that it contains loving interpersonal relationships between free beings.

---

[243] Note that we might want to argue that agents who are forced to be good, are no longer free in any sense. However this would be a considerable undertaking, and it is not necessary for us to do so to answer our question in this section. Even if we did argue this, the points following would still need to be made, and since they are simpler to establish, it seems the preferred strategy.

[244] There is a question in the wings about whether God creating is necessary or accidental. But even if creating is necessary, it need not be because there is some good that God lacks without creation, it may just be good to instantiate more of that goodness, or to share it around. This is a complex issue, but at least this indicates how we may begin to approach the question. See Morriston (2001).

[245] The is a particularly social formulation of the trinity, but Latin Trinitarians could provide an alternative description here, so not much hangs on it. See Rea (2009) and Sandler & Issler (2009).

So does the possibility of this good, that is, of a valuable freedom, require the possibility that free creatures sin? If it does, then we have a clear reason why God cannot create creatures who cannot sin without frustrating the creation of something good, which would be irrational for God, since his reason for creating is to make something good. The explanation of why valuable freedom requires the ability to sin draws on what must be the case for an agent to act for a reason. Having a character oriented towards the good, for goodness's sake, requires the exercise of freedom of choice. We noted in chapter one that if the active exercise of my ability to interact with something good is to have value, I must be responding to the goodness of that thing as I directly appreciate it. There may be a value in a robot that is programmed to always 'desire' the good, or make good movements of their bodies. But what we are focussed on here is the value there is in someone's actually valuing the goodness of something, of their loving the good, choosing the good, and being attracted to the goodness of some end. This means that I cannot be caused to desire the good, or to value good ends, but must exercise my own aseity in orienting myself regarding the good. Therefore God can create agents who have the capacity to value good ends and desire good ends. But if agents do not also have the capacity to fail to value good ends, and desire good ends, then agents have never had access to an alternative possibility concerning how they orient themselves concerning the good. If this is the case, then they never possess freedom of choice concerning this aspect of their self-formation. Now, it is not that agents are presented, early in their existence, with a single choice to be good or bad.[246] But if the formation of a self, a person, is to possess value, then any goodness that person instantiates must have been formed through the exercise of the agents aseity, or be derivative on the exercise of their aseity concerning some good-relating choice.[247] Therefore there must have been, in the history of a valuable good person, some exercise (at least one) of freedom of choice that admitted of alternative possibilities that were not all towards good ends. But this means that such agents must have had the opportunity, and ability, to sin. To create an agent without this alternative possibility would be to prevent the possibility of that agent instantiating any valuable internal relationship to the good, or active expression involving the good. And plausibly, to do so would be a bad thing to do. So this provides us with an explanation of why God could not prevent sin in the beginning.

Nagasawa, Oppy and Trakakis (2004) have offered an example that is supposed to demonstrate that agents can be created with valuable desires and states already in place. They ask us to consider John and Mary, who are a couple

[246] Matters may be different for angels, but we are focused on human agents here.
[247] And, given the prominence in agents' conception of the good in practical rationality, such choices will not be difficult to find.

in love, who are duplicated. Nagasawa et al. ask us to consider whether we would count the duplicates of John and Mary as being in love, and conclude that we should. But they note that for the duplicate versions of John and Mary there is no prior exercise of freedom, and thus conclude that it cannot be a necessary condition. However, we need to carefully consider just what has been duplicated. The duplicates would need to possess all of the memories of John and Mary, every belief and desire, every intention, and memory of every past action and intention. Just the isolated belief 'I am in love with John' would not be able to form the basis of a continuing flourishing relationship. In the John and Mary case free choices were made which produced the states necessary for relationship. They were made by the original John and Mary. The issue then is if the original John and Mary can create the right states in the duplicates, God could do so without even involving the original John and Mary – perhaps the original John and Mary do not even exist. If this were the case we would have good reason to suppose that God could create impeccable agents without the liability that libertarian freedom entails. But before the question of whether this is metaphysically possible even becomes an issue, we have to first consider whether so doing would be ethical. If not then the possibility, though interesting, is moot, since God cannot do something that is unethical. Suppose we could invent a machine that would indeed duplicate John and Mary and we are on the ethics committee considering whether the machine should be used to duplicate free agents. It would clearly not be a simple matter to decide that the machine could be so used. What if the John and Mary duplicates discovered that they were duplicates – would it destroy their love? I think that it would at least cause them to stop, and reassess their relationship.[248] But such a reassessment would involve the duplicates in an exercise of their freedom of choice regarding whether they loved each other. The fact that they would question their relationship demonstrates that such a love was not theirs to begin with. As long as they were not aware that they were duplicates, they would think they were in love, and that someone was in love with them. But the love would not be real, and would not instantiate the value that a true loving relationship would. Furthermore, there is reason to suppose that we would not be unaware of the ways in which God interacts with us in heaven, as a result of the beatific vision, so ignorance would not help keep the saints from questioning their relationships. So this example does not constitute a counter-example to our claims.

[248] The number of films exploring this issue is indicative of an intuition that duplicates do not share all of the properties of their originals. This is not an argument for this claim, clearly, but demonstrates that the claim cannot be denied without some serious thought.

Is there significant freedom in heaven?

It has been thought by some that once the self has been sufficiently oriented towards God, the alternative possibilities that remain for agents in heaven would not be ones that are significant.[249] This would mean that although we may have significant freedom on Earth, we do not have significant freedom in heaven, since we are not free to sin in heaven. We said (in chapter one) that a choice is significant if it engages an agent's capacity for freedom between alternatives that engage the agent's evaluative faculties in an assessment that recognises more than one good option, and if the choice matters to the agent because in making that choice they are determining what goods they want to instantiate, associate themselves with, and to prioritise – the choice reflects their power to not only react to the goods around them, but to exercise their freedom over who they are, and the goods that constitute their character, interests, loves, and so on. So once the redeemed have been glorified, are there any significant alternative possibilities for them to exercise their freedom between, any possibilities that enhance the value of possessing heavenly freedom?[250] Or is the only scope for the exercise of heavenly freedom rather mundane, deriving all of its value and significance derivatively?

Pawl and Timpe (2009) attempt to explain how there could be significant freedom in heaven by arguing that although we do not have the freedom to perform, or not perform, right actions, we do have the freedom to perform, or not, supererogatory actions. They explain that a right action would be to adopt the Aristotelian mean and thus do the right thing. The supererogatory aspect would be the extent to which the agent clings to the mean, the strength of their dedication to adopting the mean. Cowan (2011) has attacked this suggestion arguing that in heaven all of the redeemed will want to cling to the good, which is God. Therefore none of the redeemed will fail to perform any supererogatory action, and so it will not be a significant freedom for heavenly agents. In heaven the redeemed do not just happen to do the right thing, they are necessitated to do so. But they are necessitated in a way that involves their own psychological states. So if an agent's psychological states necessitate their adopting the mean, then it is difficult to see how they could cling any tighter to the mean. No heavenly agent will be any less clung to the mean, and no heavenly agent could possibly be more clung to the mean.

Supererogation is an attempt to provide significant alternative possibilities in heaven. But it looks like it fails to do so. Pawl and Timpe think that preventing sinful actions could still leave some open options, but worry that if the only

[249] See Rasmussen (2013), and also notes in the discussion in chapter one.
[250] We discussed why it might be though good to have certain kinds of freedom in heaven in chapter one, and we will not reiterate here.

choices are as trivial as whether to sing in the heavenly choir or to play the harp, then this would not be a significant freedom. But why look to supererogation? Pawl and Timpe do so because they constrain the sorts of possibilities that could be weighty to ones that are morally relevant. But I think we should question whether morally relevant, or significant, choices are the only non-trivial choices that will make heavenly freedom a weighty matter. If we can present an example of a heavenly choice that is open, and significant, then we will have demonstrated that Pawl and Timpe's principle is incorrect, and will hopefully have a clearer idea about what sorts of significant choices heavenly agents might explore.

Although the choice between playing the harp and singing in the choir might appear to be a trivial one, the choice of what the saint might sing to God about in heaven is not obviously trivial. Perhaps it is a requirement that saints exercise their capacities in the worship of God, but how they express themselves, and what they express themselves about, could vary. One saint might worship God by engaging in an exploration of the complexity and beauty of the cosmos, and in expressing to God their wonder at God's handiwork, and what it might reveal to them of God's character. Another saint might worship God by engaging Jesus in conversation about the experience of forgiving his executioners while on the cross, and marvelling at the depth of God's mercy and the perfection of his moral character. So we have two forms of worship, cosmic, and moral. I have chosen these two because it is very difficult to see how a human agent could engage in an appreciation of both cosmic and moral matters, at the same time. Given our cognitive limitations, the explorations of these matters seem to be distinct endeavours, and not ones that can be undertaken simultaneously, or at least not to the same degree. Since these are two forms of worship that are mutually exclusive, is the choice of which one to perform one that is open to agents? Well, given the constitution of an agent, it may be that they are particularly well-suited to a cosmic exploration rather than a moral one and that this would engage them in love of God to a greater extent. So there is some prima facie reason to think that given the constitution of a saint, there may be one choice here that is better for them. By saying that the best choice would engage their love for God in a greater way we have, plausibly, made the choice one that is a moral matter, since failing to love God, or express your love for God, could be a failure to make of the saint's heavenly relationship with God all that can be made of it. Any lapse in the realisation of the potential for good would be cause to worry that a sin had been committed.

Now agents admit of variety when they enter heaven.[251] After all, we are different from one another on Earth. There does not seem to be any reason to think that it is necessary that all the saints be made identical to each other upon glorification. After all, there is variety among the form and roles of the angels in glory. So there can at least be some variety among the saints concerning what form their worship takes. How you worship is partly constitutive of how you relate to and interact with God, therefore this variety means that there can be variety among the saints' relationships with God. But the presence of this variety does not mean that saints possess a choice. Given their constitution upon becoming a saint, the evolution of their relationship with God, the different forms of worship they engage in at different times may still be something that they are always constrained in, on pain of sinning (or failing to realise the best relationship they can), and do not possess alternative possibilities concerning. If this is the case then although there is significant variety in heaven, we are no closer to demonstrating that there can be significant freedom.

But, we need to remember that as well as agents possessing the power to apply their character to how they act, how they act also has a reciprocal effect on how they are constituted. In one way, this is quite a basic claim to make; after all, hopefully time spent discussing the cross with Jesus would effect a change in the worshipper. But we mean more than this. When agents act for a reason, that reason, the end at which they aim, the desires, beliefs and other psychological states relevant to acting for that reason, can all become more prominent in the constitution of that agent, or connected to their other states in different ways, or closer to the surface of the attention of the agent. So the effect of acting a certain way on the agent is not only whatever effect is a result of the action, but also there is an effect of having acted for that reason, or in that way, quite apart from the results that follow later. This does change the situation concerning how agents should assess whether they will worship morally, or cosmically. For a given saint, they may possess a greater potential for realising great good in their relationship with God if they worship morally rather than cosmically right now. But the measurement of the potential for great good does not only include an assessment of the good that will be realised in the performance of that action. Suppose the saint instead decided to worship God cosmically, the good realised by the act itself would be less (by definition) since the saint would be less able to express their love for God. However, by so acting, the saint would engage in self-formation that alters how they are constituted and thus alters the potential for future realisation of good. The good that can be realised by the altered saint who worships

[251] And in many theological traditions, while in heaven as well, often in significant ways, for example Stump (2012) claims that the saints can be in a greater or lesser union with God.

cosmically, even though this is not a 'strength', will be different from the good that can be realised by the saint who goes with their strength and worships morally. In the long term, there is nothing to say that the goods the cosmic worshipping saint will be able to explore, instantiate and participate in, will be of a lesser quality than if they had worshipped morally instead. The goods will be different, and the way the saint relates to them will be different, because the saint changes as well.

Let us illustrate to try and show that this is not an unusual idea. Consider a normal human agent who wants to explore beauty as expressed in suggestion. They are an artist, and so will greatly flourish in this endeavour if they study the works of the great impressionists. If they study the music of Debussy and the other impressionist composers, they will gain much less. However, in the long term, if the artist sticks at his task, he may discover more through his slowly getting to grips with music, especially as he can now relate it to his knowledge of art. Although he may not have as deep an appreciation for some details of the art, there may be details he would not have picked up on without a knowledge of music. Whichever choice the artist makes, he will be able to explore beauty in a great way, but differently. There is nothing about the fact that he is an artist that means that one choice will be worth more or less than the other. Because we are limited in our time and opportunities on earth, we are used to making decisions like this by also factoring in the practicality of what is achievable given the time we have available. However in heaven, there are no such limitations – time is on our side.

So the constitution of agents does not determine that one course of action will be one that realises the greatest potential for great good. There can be more than one possibility whose potential is not calculable in this way, since the potential will evolve and change with the agent – it may at best be a calculation that is ill-defined. This means that the choice of whether to worship cosmically or morally is one that could be open to the saint, where concerns about doing their best, realising the greatest good, and so on, do not constrain the choices of agents to a single option. This is because the choice is not simply about what the agent will do, it is also about who an agent will become. But is an open choice such as this, an alternative possibility of this type, one that makes for a significant exercise of freedom? A choice is significant if it engages the agent in evaluating and choosing between more than one good option, and if the choice matters to the agent because in making that choice they are determining what goods they want to instantiate, associate themselves with, and to prioritise – the choice reflects their power to not only react to the goods around them, but to exercise their freedom over who they are, and the goods that constitute their character, interests, loves, and so on. Well, our imagined saint will exercise themself in evaluating the options, and we have described two good options that the agent is free to choose

between. So the choice engages the saint in a non-trivial way. Also, by making their choice the agent decides what sort of person they want to be, what interest they want to pursue, what ends they want to prioritise and what goods they want to explore. This is the power of self-formation, and critically the formation of the self in relation to non-trivial matters.

Consider also that the saint's relationship with God (which may be the most significant or all matters) is partly constituted by the form of interaction that composes that relationship. So this choice is one that gives the saint power over how they relate to God. If our argument for these types of variety is correct, then the beatific vision is not the instantiation of a perfect relationship with God, because there is no single concept of a perfect relationship with God.[252] Certainly the relationship is developed so that it does not contain any flaw, so there are no false beliefs about God. But, if we think carefully, it is not surprising that there is no single perfect relationship with God, for if by perfect relationship we mean one that instantiates all true beliefs about God and explores all forms of interaction involving those beliefs, then such a relationship is impossible for a creature to possess. Creatures are finite, and limited in their cognitive capacities; thus, they are not able to apprehend all of God. So there is a limitation on the saints' relationships with God. However, the saints are free to explore how they relate to God and what can be known of God. It is just that this may be an exploration that is without end. So the beatific vision does not impart a perfect relationship to the saint; however, it does perfect the saint regarding their being equipped to enjoy God and to grow in their relationship with God.[253] How this relationship grows, and what facets of God are explored when, and how, are matters that the agent has freedom of choice over, though God will provide input to the direction of the relationship too (after all it is not a relationship with something inanimate, but an active, three-personned Godhead). Therefore, saints do possess alternative possibilities that are significant, and there is no need for them to be able to sin to possess such possibilities.

[252] Perhaps God may instantiate perfect relationship, but this is because God does not change, and cannot change.

[253] If the analogy helps, being perfectly proficient in the capacities required to play the violin would not stop me from developing in my enjoyment of playing a piece of music. Furthermore, as I play, it engages me emotionally, and changes me, so that if I play it again, the experience has some new dimension. The analogy is limited as I may, perhaps, be able to exhaust the possibilities of engaging with a piece of music, whereas God is limitless, and the prospects for growth are richer.

Conclusion

So, we have explored the conditions under which agents possess freedom. We considered how agents can be the sources of what they do when they are able to align themselves, and what they do, with ends that they value as good, when they possess reasons for acting in light of their sensitivity to their place in the world, and when they can guide their action in light of these concerns. However free agents also need to possess some causal independence from the past, so we must be incompatibilists about free agency. This means that agents must have access to some alternative possibilities. We defended a non-causal theory of control that showed that it is coherent to claim that agents possess genuine control over which actions they perform. This independence and control secures properties we need agents to possess for the free will defence, so they can be morally and agentially responsible for what they do.

We discovered that given the nature of this capacity for freedom, and its role in instantiating great goods, it is impossible for God to prevent sin unless agents exercise their agency in a particular way first – in submitting to God. The space between the acquisition of freedom and this submission, marks the space in which sin is possible for agents but God is not able to prevent them sinning. This explains why God cannot prevent sin, without destroying freedom. However, there is hope for free agents, for once agents have exercised their freedom in submission, God is able to prevent future sin in heaven. Moreover, God's prevention of sin in heaven does not leave agents without significant freedom, and there is scope for the exercise of heavenly freedom to involve agents in significant freedom of choice, and in control over who they are, their relationships, what they do – which are aspects of our agency we care greatly about.

Therefore, our two claims

iii) God cannot prevent sin without destroying free will;
iv) In heaven, God prevents sin without destroying free will.

are not contradictory, and we have good reason to maintain both claims as not only consistent with each other (which was our primary aim), but also, hopefully as well-grounded in our experience of agency, and a plausible analysis of what freedom actually does consists in.

References

Alvarez, M. (2001). Letting happen, omissions and causation. *Grazer Philosophische Studien*, 61, 63-82.

Alvarez, M. (2009a). Acting intentionally and acting for a reason. *Inquiry*, 52(3), 293-305.

Alvarez, M. (2009b). Actions, thought-experiments and the 'principle of alternate possibilities'. *Australasian Journal of Philosophy*, 87(1), 61-81.

Alvarez, M. (2010). *Kinds of Reasons*. Oxford: Oxford University Press.

Alvarez, M. (2013). Agency and Two-Way Powers. *Proceedings of the Aristotelian Society,* 113, 101–121.

Alvarez, M. & Hyman, J. (1998). Agents and their Actions. *Philosophy*, 73(2), 219–245.

Anscombe, G. E. M. (1957). *Intention*. Harvard University Press.

Anscombe, G. E. M. (1971). *Causality and Determinism*, Cambridge: Cambridge University Press.

Anselm (1998). *Anselm of Canterbury: The Major Works*. Oxford: Oxford University Press.

Antony, L. M. (2007). *Philosophers without Gods: Meditations on Atheism and the Secular Life*. Oxford: Oxford University Press.

Arpaly, N. (2010). On Acting Rationally against One's Best Judgment*. *Ethics,* 110(3), 488–513.

Austin, J. L. (1956). Ifs and cans. In Austin, J. (1961). *Philosophical papers*, ed. J. Urmson and G. Warnock, 153–180. Oxford: Clarendon Press.

Bach, K. (2010) Refraining, omitting, and negative acts. In T. O'Connor & C. Sandis, eds. *A Companion to the Philosophy of Action*. Malden, MA: Blackwell.

Baumeister, R. F., Mele, A. R. & Vohs, K., eds. (2010). *Consciousness and free action: how might they work?* New York: Oxford University Press.

Beebee, H. & Mele, A. (2002). Humean compatibilism. *Mind*, 111(442), 201-224.

Bratman, M. (1987). *Intention, Plans, and Practical Reason*. Cambridge, Mass.: Harvard University Press.

Bratman, M. E. (1997). Responsibility and Planning. *The Journal of Ethics*, 1(1), 27–43.

Bratman, M. E. (2000). Valuing and the Will. *Noûs*, 34(14), 249–265.

Bratman, M. E. (2003). A desire of one's own. *The Journal of Philosophy*, 100(5),

221-242.

Bratman, M. E. (2004). Three Theories of Self-Governance. *Philosophical Topics*, 32 (1), 21–46.

Bratman, M. (2008). Temptation revisited. In B. Verbeek, ed. *Reasons and Intentions*. Ashgate.

Chisholm, R. (1964a). *Human freedom and the self. The Langley Lecture, 1964.* Repreinted in G. Watson, ed. (2003). *Free Will*. Oxford: Oxford University Press.

Chisholm, R. M. (1964b). The Descriptive Element in the Concept of Action. *The Journal of Philosophy*, 61(20), 613-625.

Clarke, R. (1993). Toward a credible agent-causal account of free will. *Noûs*, 27(2), 191-203. Reprinted in T. O'Connor ed. (1995). *Agents, Causes, and Events*. Oxford University Press.

Clarke, R. (2000). Modest Libertarianism. *Noûs*, 34 (14), 21–45.

Clarke, R. (2003). *Libertarian Accounts of Free Will*. Oxford: Oxford University Press.

Cowan, S. B. (2011). Compatibilism and the Sinlessness of the Redeemed in Heaven. *Faith and Philosophy*, 28(4), 416–431.

Craig, W. L. (1989). "No Other Name". *Faith and Philosophy*, 6(2), 172-188.

Crisp, O. D. (2004). Did Christ have a Fallen Human Nature? *International Journal of Systematic Theology*, 6(3), 270–288.

Crisp, O. D. (2007a). *Divinity and Humanity*. Cambridge: Cambridge University Press.

Crisp, O. D. (2007b). Was Christ Sinless or Impeccable? *Irish Theological Quarterly*, 72(2), 168–186.

Crisp, O. D. (2009). *God Incarnate*. London: T&T Clark.

Crisp, O. D. (2011). Analytic Theology. *The Expository Times*, 122(10), 469–477.

Crisp, O. D. & Rea, M. C., eds. (2009). *Analytic Theology: New Essays in the Philosophy of Theology*. Oxford: Oxford University Press.

D'Oro, G. & Sandis, C., eds. (2013). *Reasons and Causes*. Basingstoke: Palgrave Macmillan.

Dancy, J. (1995). Why There Is Really No Such Thing as the Theory of Motivation. *Proceedings of the Aristotelian Society*, 95, 1-18.

Dancy, J. (2002). *Practical Reality*. Oxford: Oxford University Press.

Dancy, J. (2004). Two Ways of Explaining Actions. *Royal Institute of Philosophy*

Supplement, 55, 25–42.

Danker, F. W. (2000). *A Greek-English Lexicon of the New Testament and Other Early Christian Literature (Bauer, BDAG)*. 3rd edition. Chicago: The University of Chicago Press.

Danto, A. C. (1965). Basic Actions. *American Philosophical Quarterly*, 2(2), 141-148.

Davidson, D. (1963). Actions, Reasons, and Causes. *The Journal of Philosophy*, 60(23), 685-700.

Davidson, D. (1969). How is weakness of the will possible? In D. Davidson (1980). *Essays on actions and events.* Oxford: Clarendon Press.

Davidson, D. (1971). Agency. In D. Davidson (1980). *Essays on actions and events.* Oxford: Clarendon Press.

Davidson, D. (1973) Freedom to act. In D. Davidson (1980). *Essays on actions and events.* Oxford: Clarendon Press.

Davidson, D. (1980). *Essays on actions and events.* Oxford: Clarendon Press.

Davis, W. A. (2005). Reasons and psychological causes. *Philosophical Studies*, 122(1), 51–101.

Demarest, B. (2006). *The Cross and Salvation*. Wheaton, IL: Crossway.

Dennett, D. C. (1984). *Elbow Room*. Oxford: Oxford University Press.

Donnelly, J. (1985). Eschatological enquiry. *Sophia*, 24(3), 16–31.

Dretske, F. I. (1988). *Explaining Behavior*. Cambridge, MA: MIT Press.

Dretske, F. (2010). Triggering and Structuring Casues. In T. O'Connor & C. Sandis, eds. *A Companion to the Philosophy of Action*. Malden, MA: Blackwell.

Edwards, J. (1754). *Freedom of the Will*. A. S. Kaufman and W. K. Frankena, eds. Bobbs-Merrill (this edition 1969).

Ekstrom, L. W. (1993). A coherence theory of autonomy. *Philosophy and Phenomenological Research*, 53(3), 599-616.

Ekstrom, L. W. (2000). *Free Will*. Boulder, CO: Westview.

Ekstrom, L. W. (2005). Alienation, Autonomy, and the Self. *Midwest Studies in Philosophy*, 29(1), 45–67.

Elwell, W. A. (2001). *Evangelical Dictionary of Theology*. Grand Rapids, MI: Baker Academic.

Eshleman, A. (2014). Moral responsibility. In *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta ed. [online].

Everson, S. (2010). Motivating Reasons. In T. O'Connor & C. Sandis, eds. *A Companion to the Philosophy of Action*. Malden, MA: Blackwell.

Fischer, J. M. (1993). *The Metaphysics of Death*. Stanford, CA: Stanford University Press.

Fischer, J. M. (1994). *The Metaphysics of Free Will*. Malden, MA: Wiley-Blackwell.

Fischer, J. M. (1999). Recent Work on Moral Responsibility*. *Ethics*, 110(1), 93-139.

Fischer, J. M. (2012). *Deep Control*. New York: Oxford University Press.

Fischer, J. M. & Ravizza, M. (1998). *Responsibility and Control*. Cambridge: Cambridge University Press.

Fischer, J. M. et al. (2007). *Four Views on Free Will*. Malden, MA: John Wiley & Sons.

Frankfurt, H. G. (1971). Freedom of the Will and the Concept of a Person. *The Journal of Philosophy*, 68(1), 5–20.

Frankfurt, H. G. (1978). The problem of action. *American Philosophical Quarterly*, 15(2), 157-162.

Frankfurt, H. G. (1987). Identification and Wholeheartedness. In F. Schoeman, ed. *Responsibility, Character, and the Emotions*. Cambridge: Cambridge University Press.

Frankfurt, H. G. (1994). Autonomy, Necessity, and Love. In Fulda and Horstman eds. *Vernunftbegriffe in der Moderne*. Stuttgart: Stuttgart Hegel-Kongress

Gaine, S. (2003). *Will There Be Free Will in Heaven?* London: T & T Clark.

Geisler, N. (2004). *Systematic Theology: Sin and Salvation*. Minneapolis, MI: Bethany House.

Gibb, S. C. et al., eds. (2013). *Mental Causation and Ontology*. Oxford: Oxford University Press.

Ginet, C. (1990). *On Action*. Cambridge: Cambridge University Press.

Ginet, C. (1995). Reason's explanation of action. In T. O'Connor, ed. *Agents, Causes, and Events*. New York: Oxford University Press.

Goetz, S. (2008). *Freedom, Teleology, and Evil*. London: Continuum.

Grudem, W. A. (1994). *Systematic Theology*. Grand Rapids, MI: IVP.

Haji, I. (2012). *Reason's Debt to Freedom*. Oxford: Oxford University Press.

Heil, J. & Mele, A., eds. (1995). *Mental Causation*. Oxford: Oxford University Press.

Hodgson, D. (2012). *Rationality + Consciousness = Free Will*. Oxford: Oxford University Press.

Hoefer, C. (2010). Causal Determinism. In *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta ed. [online].

Holton, R. (1999). Intention and weakness of will. *The Journal of Philosophy*, 96(5), 241-262.

Hornsby, J. (1980). *Actions*. London: Routledge.

Hornsby, J. (2004). Agency and Actions. *Royal Institute of Philosophy Supplement*, 55, 1–23.

Howard-Snyder, D. & Moser, P. K., eds. (2002). *Divine Hiddenness*. Cambridge: Cambridge University Press.

Hume (1748). *An Enquiry concerning Human Understanding*. S. Buckle, ed. Cambridge University Press (this edition 2007).

Hursthouse, R. (1991). Arational actions. *Journal of Philosophy*, 88(2), 57-68.

Hutto, D. (2013). Still a cause for concern: reasons, causes and explanations. In G. D'Oro & C. Sandis, eds. *Reasons and Causes*. Basingstoke: Palgrave Macmillan.

Hyman, J. (2006). Three fallacies about action. *Proceedings of the 29th International Wittgenstein Symposium*. Hölder-Pichler-Tempsky.

Hyman, J. & Steward, H. eds. (2004). *Agency and Action*. Cambridge: Cambridge University Press.

Kane, R. (1998). *The Significance of Free Will*. New York: Oxford University Press.

Kane, R. (2002a). *The Oxford Handbook of Free Will*, First ed. Oxford: Oxford University Press.

Kane, R., ed. (2002b) *Free Will*. Malden, MA: Wiley-Blackwell.

Kane, R. (2011). Rethinking Free Will: New Perspectives on an Ancient Problem. In R. Kane, ed. *The Oxford Handbook of Free Will*. Oxford: Oxford University Press.

Kane, R., ed. (2011). *The Oxford Handbook of Free Will*. Second Ed. Oxford: Oxford University Press.

Kreeft, P. (2001). *Catholic Christianity*. San Francisco: Ignatius Press.

Kvanvig, J. L., ed. (2014). *Oxford Studies in Philosophy of Religion, Vol. 5*. Oxford: Oxford University Press.

Lehrer, K. (1964). 'Could'and Determinism. *Analysis*, 24(4), 159-160.

Lewis, D. (1981). Are we free to break the laws? *Theoria*, 47(3), 113–121.

Levy, N. (2011). *Hard luck: how luck undermines free will and moral responsibility*. New York: Oxford University Press.

Levy, N. (2012). Consciousness, Implicit Attitudes and Moral Responsibility. *Noûs*, 48(1), 21-40.

Levy, N. (2014). *Consciousness and Moral Responsibility*. Oxford: Oxford University Press.

Lowe, E. J. (2001). Event causation and agent causation. *Grazer Philosophische Studien*, 61, 1-20.

Lowe, E. J. (2010). *Personal Agency*. Oxford: Oxford University Press.

McCann, H. (1974). Volition and Basic Action. *The Philosophical Review*, 83(4), 451.

McCann, H. (1998). *The Works of Agency*. Ithaca: Cornell University Press.

McGrath, S. (2005). Causation By Omission: A Dilemma. *Philosophical Studies*, 123(1-2), 125–148.

McLaughlin, B. (2013). Why Rationalization is Not a Species of Causal Explanation. In G. D'Oro & C. Sandis, eds. *Reasons and Causes*. Basingstoke: Palgrave Macmillan.

Markosian, N. (1999). A compatibilist version of the theory of agent causation. *Pacific Philosophical Quarterly*, 80(3), 257-277.

Melden, A. I. (1961). *Free Action*. London: Routledge & Kegan Paul.

Mele, A. R. (1992). *Springs of Action: Understanding Intentional Behavior*. New York: Oxford University Press.

Mele, A. R. (1995). *Autonomous Agents: From Self-Control to Autonomy*. Oxford: Oxford University Press.

Mele, A. R. (1997). *The Philosophy of Action*. Oxford: Oxford University Press.

Mele, A. R. (2000). Goal-Directed Action: Teleological Explanations, Causal Theories, and Deviance. *Noûs*, 34(14), 279-300.

Mele, A. R. (2003). Agents' abilities. *Noûs*, 37(3), 447-470.

Mele, A. R. (2004). Can Libertarians Make Promises? *Royal Institute of Philosophy Supplement*, 55, 217–241.

Mele, A. R. (2006). *Free Will and Luck*. Oxford: Oxford University Press.

Mele, A. (2010). Motivational Strength. In T. O'Connor & C. Sandis, eds. *A Companion to the Philosophy of Action*. Malden, MA: Blackwell.

Morriston, W. (2001). Omnipotence and necessary moral perfection: are they compatible? *Religious Studies*, 37(2), 143–160.

Müller, T. (2012). Branching in the landscape of possibilities. *Synthese*, 188(1), 41-65.

Muraven, M. & Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological bulletin*, 126(2), 247-259.

Murphy, N. C. & Brown, W. S. (2009). *Did my neurons make me do it? Philosophical and neurobiological perspectives on moral responsibility and free will*. New York: Oxford University Press.

Murphy, N., Ellis, G. & O'Connor, T., eds. (2009). *Downward Causation and the Neurobiology of Free Will*. Berlin: Springer.

Nagasawa, Y., Oppy, G. & Trakakis, N. (2004). Salvation in heaven? *Philosophical papers*, 31(1), 97-119.

Nozick, R. (1981). *Philosophical Explanations*. Cambridge, MA: Harvard University Press.

O'Collins, G. (2009). *Christology*. Oxford: Oxford University Press.

O'Connor, T. (1995). *Agents, Causes, and Events*. Oxford: Oxford University Press.

O'Connor, T. (2000). *Persons and Causes*. New York: Oxford University Press.

O'Connor, T. (2009). Degrees of freedom. *Philosophical Explorations*, 12(2), 119-125.

O'Connor, T. (2010). Free Will. In *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta ed. [online].

O'Connor, T. & Sandis, C. (2010) *A Companion to the Philosophy of Action*. Malden, MA: Blackwell.

O'Shaughnessy (2008). *The Will*. Cambridge: Cambridge University Press.

Pawl, T. & Timpe, K. (2013). Heavenly Freedom: A Response to Cowan. *Faith and Philosophy*, 30(2), 188-197.

Pawl, T. & Timpe, K. (2009). Incompatibilism, Sin, And Free Will In Heaven. *Faith and Philosophy*, 26(4), 398–419.

Pereboom, D. (2001). *Living without Free Will*. Cambridge: Cambridge University Press.

Perszyk, K., ed. (2011). *Molinism*. Oxford: Oxford University Press.

Pickard, H. (2013). Psychopathology and the Ability to Do Otherwise. *Philosophy and Phenomenological Research*, [online].

Pink, T. (1996). *The Psychology of Freedom*. Cambridge: Cambridge University Press.

Pink, T. (2004). *Free Will: A Very Short Introduction*. Oxford: Oxford University Press.

Pink, T. (2008). Intentions and two models of human action. In B. Verbeek, ed. *Reasons and intentions*. Aldershot: Ashgate.

Pink, T. (2011). Freedom and Action without Causation: Noncausal Theories of Freedom and Purposive Agency. In R. Kane, ed. *The Oxford Handbook of Free Will, Second Ed*. Oxford: Oxford University Press.

Pink, T. & Stone, M. W. F., eds. (2004). *The Will and Human Action, from Antiquity to the Present Day*. London: Routledge.

Plantinga, A. (1974). *God, Freedom, and Evil*. Grand Rapids, MI: Wm. B. Eerdmans Publishing.

Pollard, B. (2003). Can virtuous actions be both habitual and rational? *Ethical Theory and Moral Practice*, 6(4), 411-425.

Pollard, B. (2006). Explaining actions with habits. *American Philosophical Quarterly*, 43(1), 57-69.

Pollard, B. (2010). Habitual Actions. In T. O'Connor & C. Sandis, eds. *A Companion to the Philosophy of Action*. Malden, MA: Blackwell.

Rasmussen, J. (2013). On the Value of Freedom To Do Evil. *Faith and Philosophy*, 30(4), 418–428.

Rea, M. C. (2009). *Oxford Readings in Philosophical Theology: Volume 1*. Oxford: Oxford University Press.

Reid, T. (1843). *Essays on the Active Powers of the Human Mind*. D. Stewart & G. N. Wright, eds. Forgotten Books (this edition 2012).

Rogers, K. (2008). *Anselm on Freedom*. New York: Oxford University Press.

Ruben, D. (1999). Act individuation: the Cambridge theory. *Analysis*, 59(264), 276-283.

Ruben, D (2003). *Action and Its Explanation*. Oxford: Oxford University Press.

Sanders, F. & Issler, K. (2007). *Jesus in Trinitarian Perspective*. Nashville, TN: B&H Publishing Group.

Sandis, C. (2008). How to Act Against Your Better Judgement. *Philosophical Frontiers: A Journal of Emerging Thought*, 3(2), 111-124.

Sandis, C. (2012a). *The Things We Do and Why We Do Them*. Basingstoke: Palgrave MacMillan.

Sandis, C. (2012b). The Objects of Action Explanation. *Ratio*, 25(3), 326–344.

Schueler, G. F. (1995). *Desire*. Cambridge, MA: MIT Press.

Schueler, G. F. (2003). *Reasons and Purposes : Human Rationality and the Teleological Explanation of Action*. Oxford: Oxford University Press.

Schueler, G. F. (2009). The Humean Theory of Motivation Rejected1. *Philosophy and Phenomenological Research*, 78(1), 103–122.

Schroeder, T. (2010). Desire and Pleasure. In T. O'Connor & C. Sandis, eds. *A Companion to the Philosophy of Action*. Malden, MA: Blackwell.

Sehon, S. R. (1997). Deviant Causal Chains and the Irreducibility of Teleological Explanation. *Pacific Philosophical Quarterly*, 78(2), 195–213.

Sehon, S. R. (2000). An argument against the causal theory of action explanation. *Philosophy and Phenomenological Research*, 60(1), 67-85.

Sehon, S. R. (2005). *Teleological realism*. Cambridge, MA: MIT Press.

Sehon, S (2010). Teleological Explanation. In T. O'Connor & C. Sandis, eds. *A Companion to the Philosophy of Action*. Malden, MA: Blackwell.

Sennett, J. F. (1999). Is There Freedom In Heaven? *Faith and Philosophy*, 16(1), 69–82.

Setiya, K. (2014). Intention. In *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta ed. [online].

Smith, M. (1987). The Humean theory of motivation. *Mind*, XCVI(381), 36-61.

Steel, C. (2004). The effect of the will on judgement: Thomas Aquinas on faith and prudence. In T. Pink & M. W. F. Stone, eds. *The Will and Human Action, from Antiquity to the Present Day*. London: Routledge.

Steward, H. (1997a). On the Notion of Cause'Philosophically Speaking'. *Proceedings of the Aristotelian Society*, 97, 125-140.

Steward, H. (1997b). *The Ontology of Mind*. Oxford: Oxford University Press.

Steward, H. (2000). Do actions occur inside the body? *Mind & Society*, 1(2), 107–125.

Steward, H. (2009). The truth in compatibilism and the truth of libertarianism. *Philosophical Explorations*, 12(2), 167–179.

Steward, H. (2012). *A Metaphysics for Freedom*. Oxford: Oxford University Press.

Stout, R. (2005). *Action*. Chesham: Acumen.

Stout, R. (2010a). Deviant Causal Chains. In T. O'Connor & C. Sandis, eds. *A Companion to the Philosophy of Action*. Malden, MA: Blackwell.

Stout, R. (2010b). What are you causing in acting? In J. H. Aguilar and A. A. Buckareff eds. *Causing Human Action: New Perspectives on the Causal Theory of Action*. Cambridge, MA: MIT Press.

Stoutland, F. (1968). Basic actions and causality. *The Journal of Philosophy*, 65(16), 467-475.

Stoutland, F. (2001) Responsive action and the belief-desire model. *Grazer Philosophische Studien*, 61, 83-106.

Strawson, P. F. (2008). *Freedom and Resentment and Other Essays*. London: Routledge.

Strawson, G. (2011). Free will. In *Routledge Encyclopedia of Philosophy*, E. Craig, ed. [online].

Strong, A. H. (1907). *Systematic Theology: A Compendium and Commonplace Book Designed for the Use of Theological Students*. London: Pickering & Inglis.

Stump, E. (2012). *Wandering in Darkness*. New York: Oxford University Press.

Taliaferro, C. & Meister, C. (2010). *The Cambridge Companion to Christian Philosophical Theology*. Cambridge: Cambridge University Press.

Thomson, J. J. (2003). Causation: Omissions. *Philosophy and Phenomenological Research*, 66(1), 81–103.

Timpe, K. (2008). *Free Will*. London: Continuum.

Timpe, K. (2014). The Arbitrariness of the Primal Sin. In J. L. Kvanvig, ed. *Oxford Studies in Philosophy of Religion, Vol. 5*. Oxford: Oxford University Press.

Van Inwagen, P. (1983). *An Essay on Free Will*. Oxford: Oxford University Press.

Van Inwagen, P. (1994). When the will is not free. *Philosophical Studies*, 75(1-2), 95-113.

Van Inwagen, P. (2000). Free Will Remains a Mystery. *Noûs*. 34(s14), 1–19.

Van Inwagen, P. (2004). Freedom to Break the Laws. *Midwest Studies in Philosophy*, 28(1), 334–350.

Van Inwagen, P. (2008). How to Think about the Problem of Free Will. *The Journal of Ethics*, 12(3-4), 327–341.

Velleman, J. D. (1992). What happens when someone acts? *Mind*, 101(403), 461-481.

Verbeek, B. (2008). *Reasons and intentions*. Aldershot: Ashgate.

Vermazen, B. (1985). Negative acts. In B. Vermazen & M. B. Hintikka, eds. *Essays on Davidson*. Oxford: Oxford University Press.

Vihvelin, K. (2013). *Causes, Laws, and Free Will*. Oxford: Oxford University Press.

Wallace, R. J. (2004). Normativity and the Will. *Royal Institute of Philosophy Supplement*, 55, 195–216.

Walls, J. L. (2002). *Heaven*. Oxford: Oxford University Press.

Walls, J. L. (2010). Heaven and hell. In C. Taliaferro & C. Meister, eds. *The Cambridge Companion to Christian Philosophical Theology*. Cambridge: Cambridge University Press.

Watson, G. (1975). Free agency. *The Journal of Philosophy*, 72(8), 205-220.

Watson, G., ed. (2003). *Free Will*. Oxford: Oxford University Press.

Wiggins, D. (1987). Towards a reasonable libertarianism. Reprinted in D. Wiggins (1998). *Needs, Values, Truth*. Oxford: Oxford University Press.

Wiggins, D. (1998). *Needs, Values, Truth*. Oxford: Oxford University Press.

Wilson, G. & Shpall, S. (2012). Action. In *The Stanford Encyclopedia of Philosophy*. Edward N. Zalta ed. [online].

Wolf, Susan (1987). Sanity and the Metaphysics of Responsibility. In F. Schoeman, ed. *Responsibility, Character, and the Emotions*. Cambridge: Cambridge University Press.

Wolf, S. (1993). *Freedom Within Reason*. Oxford: Oxford University Press.

Wolf, S. (1997). Happiness and Meaning: Two Aspects of the Good Life. *Social Philosophy and Policy*, 14(1), 207–225.