4th BEIJING ANALYTIC PHILOSOPHY CONFERENCE
18 March 2017

TIMOTHY O'CONNOR
(Indiana University)

“Chance, Probability, and the Explanation of Human Decisions”

COMMENTS

TANG REFENG
(Chinese Academy of Social Sciences)

Venue
Room 600
Renwen Building
Renmin University

Language
English

Organizers
Chen Ju
Li Kun
Liu Chang
Tian Jie
Wang Ping
Yu Junwei
Daniel Lim
CONFERENCE SCHEDULE

9:20 – 9:30  Introduction

9:30 – 10:05  *A New Indifference Argument Against Motivational Internalism* (3-13)
Zhang Wan (Capital Normal)
Comments: Zhou Yili (Renmin)

10:10 – 10:45  *Does a Physical Property Count as a Confounder?* (14-24)
Xin Dong (Peking)
Comments: Liu Chang (Renmin)

Zhao Yi (Fudan)
Comments: Timothy O'Connor (Indiana)

11:30 – 12:05  *Situationism and Moral Responsibility* (38-47)
Zhang Ke (Renmin)
Comments: Tian Jie (Renmin)

12:10  Group Photo

12:15 – 1:25  Lunch Break

1:30 – 2:45  Keynote Address

*Chance, Probability, and the Explanation of Human Decisions* (48-56)
Timothy O'Connor (Indiana)
Comments: Tang Refeng (Chinese Academy of Social Sciences)

2:50 – 3:00  Essay Prizes & Closing Comments
A New Indifference Argument against Motivational Internalism

ZHANG WAN

1. Introduction

The debate between motivational internalism and externalism is an important issue in contemporary metaethics. Motivational internalism holds that first-person singular present-tense moral judgments are conceptually necessarily connected with appropriate motivations.¹ Today, the widely accepted formulation of motivational internalism is as follows:

Simple internalism: Necessarily, if a person judges that she morally ought to φ, then she is (at least somewhat) motivated to φ (Björnsson, G., et al. 2012).

On the contrary, motivational externalism is just the negation of motivational internalism, claiming that it’s possible for an agent to make a moral judgment while no corresponding motivation occurs.² The paradigm strategy used by externalists is to appeal to indifference arguments, which consider the possibility of agents who remain indifferent to the moral judgments they make (Mele 1996; Shafer-Landau 2000; Joyce 2001; Strandberg 2004; Miller 2008). Resorting to indifference arguments is a natural strategy for externalists, since if counterexamples of agents being indifferent to moral judgments they hold are conceivable, then these cases are conceptually possible, which means externalism is true and internalism is false.

Traditional indifference arguments come in various versions. To re-characterize them briefly, they are the argument from depression and the argument from brain trauma (Shafer-Landau 2000, p. 273-4; Mele 1996, p. 73), the argument from extreme emotions (Shafer-Landau 2000, p.273; Strandberg 2004, p. 191; Miller 2008, p. 249) and the argument from a purely evil agent (Joyce 2001, p. 21).

In responding to those indifference arguments, internalists basically deploy two kinds of strategies: they either defuse those indifference arguments directly, or accept them but qualify the rendition of internalism to avoid them. The defusing strategy

¹ Let me make it clear here that ‘motivations’ are understood as causal mental states that (defeasibly) bring about intentional actions.

² For externalists, the occurrence of moral motivations must rely on the combination of moral judgments and other distinctive conative states (Shafer-Landau 2003; Zangwill 2008; Swartzer 2015).
maintains that moral motivations actually occur in some traditional indifferent cases (Park 2016, Bedke 2009), or that moral judgments without motivation are not sincere or genuine moral judgments (Bromwich 2016), or that externalists cannot account for some traditional indifference cases either (Swartzer 2015). The avoiding strategy, on the other hand, comes up with conditional internalism or weaker versions of internalism to shun traditional indifference arguments (Björnsson, G., et al. 2012, Eggers 2015).

Internalists’ recent objections have posed considerable threats to traditional indifference arguments, and to me, that’s mainly due to externalists’ excessive reliance on abnormal agents – agents who have abnormal psychological states, who are grasped by extreme emotions and who are abnormally evil – which makes it difficult to judge intuitively whether the agent has a motivation corresponding to a sincerely accepted moral judgment. It also tends to make one believe that externalists can only win individual battles (Blackburn 1998).

It’s the aim of this paper to explore a new indifference argument for externalists to better cope with the above and other potential challenges posed by internalists. The organization of this paper is as follows. In Section 2, I’ll develop a new indifference argument, the argument from reflective moral inference (ARMI), which shows that a completely normal moral agent in a normal case of moral practice can disassociate honestly accepting a moral judgment from having the related motivation. Then in Section 3, I'll take internalists’ recent objections into consideration and see whether they may threat ARMI.

2. A New Indifference Argument: The Argument from Reflective Moral Inference (ARMI)

Before we move on to the new argument, I want to introduce a new term. I think we all agree that all moral judgments can be used as premises, middle steps or the final conclusion of a series of inferences. In that sense I’ll say that moral judgments have the inferential property. As we’ll see, the argument below will be based on the inferential property of moral judgments. The important point is that the inferential property is a universal property of all moral judgments themselves, regardless of what special agents are making the moral judgments or in which special circumstances the moral judgments
are made. That is how my new argument differs from traditional indifference arguments.

Consider the case below:

Pi is a rational and normal young man who survives a sea storm and manages to get in a lifeboat alone. Several days later, another storm wipes out all the food and drink from the lifeboat. When Pi wakes up he finds a big living fish in the boat and becomes aware that all the previous food storage is gone. Pi is a vegetarian because of his religious beliefs. He knows the fish is a living creature which is edible from books, but he has never had any chance to eat fish before, nor has he ever thought about the moral status of releasing, killing or eating fish, since he was brought up in a place where no fish can be seen. Pi also knows that it’s impossible to find vegetarian food on the sea. Although a little tired and frustrated, Pi is psychologically normal and intact, as rational as he has always been, and as virtuous as any virtuous person in our society is. Now looking at the big fish, he has no idea whether to eat it or not at the start. So Pi sinks into pondering. And here is how he goes through the reasoning:

(1) Everything my religion says is true.
(2) My religion says that I ought to set captive lives free instead of killing them.
(3) I ought to set captive lives free instead of killing them.
(4) If I ought to set captive lives free instead of killing them, then I ought to set the fish free instead of killing it.
(5) I ought to set the fish free instead of killing it.
(6) If I don’t kill the fish, I’ll have nothing to eat, and thus will starve myself to death.
(7) For me, starving myself to death is a bigger evil than killing the fish.
(8) When two evils can’t be avoided at the same time, choose the lesser of the two evils.
(9) I ought to kill and eat the fish.

A successful indifference argument should meet three criteria: it should (a) involve a sincere moral judgment; (b) involve no corresponding motivation; and (c) fit into the externalist explanation. Let’s examine what happens when Pi comes to accept (5) in his inference.

First, (5) is a sincere moral judgment concerning his own situation, and it’s no doubt a first-person present-tense statement about the moral rightness of fish eating. As
we can see, it's logically implied by (3) and (4). Therefore, if Pi really accepts (3) and (4), then he is rationally committed to accepting (5). It turns out that Pi accepts (3) as the conclusion of a valid argument and accepts (4) because he knows the fish in the boat is a living being. Then as a rational and self-consistent person, Pi will certainly go on to accept (5) sincerely. Besides, Pi makes (5) when setting the fish free or not is a very practical choice for him. He is not imagining any irrelevant hypothetical situation. So this definitely belongs to the kind of sincere moral judgments that concern us in the debate.

Then what about the motivation? Does a motivation to free the fish ever occurs to Pi when he judges (5)? My answer is no. Intuitively, what Pi does in the reasoning is just to assess each of the statements. He makes the moral judgment (5) by tracing back the inferential evidences that imply (5), and then he goes on to add more statements to his inference to reach the final conclusion. His attention is focused on the inference, and he doesn’t perform any action associated with (5), like throwing the fish into the sea during his reasoning. Besides, since we don’t detect any special elements in this case that might have overridden an existent motivation, it seems implausible that the absence of action is obstructed by such an element, instead of due to the absence of motivation.

Furthermore, as Pi has never thought about the moral status of releasing, killing or eating fish before, (5) is a totally new judgment for him. There is no emotional or experiential resources he can appeal to when he judges (5). So (5) comes to him only as a statement in the pure inference. We can hardly imagine where certain motivation could come from when Pi judges (5).

Of course, internalists may insist that even if (5) is a totally new statement for Pi, and Pi appears to have no motivation in his reasoning, the moral motivation actually occurs without Pi’s notice. They may insist that the motivation dwells in the meaning of the word ‘ought’. So as long as Pi makes a first person ‘ought’ judgment, he would have the appropriate motivation to act in accordance with it. But this response, attractive as it seems, is straightly begging the question against externalism by assuming that ‘ought’

---

3 We’ve described Pi as a psychologically normal, rational and fully competent agent who is conducting a peaceful reasoning, as is opposed to traditional cases where special elements like abnormal psychological states and extreme emotions may have obstructed the putative existent motivations from generating appropriate actions.
is conceptually necessarily associated a motivation. It makes one wonder why not all ‘ought’ sentences -- including third-person and future-tense sentences – have the same capacity to trigger motivations. Besides, such a view is confronted with a Frege-Geach style objection: it couldn’t give a successful account for ‘ought’ in complex moral sentences such as conditionals (Schroeder 2008, Sander 2014).

Internalists may also claim that Pi actually has a motivation to set the fish free when he judges (5), but since he is in the process of an inference, he suspends the motivation and disassociate it from any action. The motivation to set the fish free then lingers in Pi’s mind without being noticed, until Pi makes a contrary judgment (9) and then the former motivation becomes defeated by the conflicting motivation (to kill and eat the fish). This objection seems intuitively plausible, but as we have noted in the introduction, a motivation is a mental state that (defeasibly) causally triggers some action. If a motivation fails to trigger any action, there must be some stronger conflicting mental states weighing out that motivation and obstruct the connection between that motivation and the appropriate action. However, in our case, Pi is peacefully conducting the reflective reasoning. He doesn’t have any strong conflicting mental states and his psychological make-up is sound and normal. How could the putative motivation become suspended or overridden?

Then internalists may further contend: could there be a pre-existing desire-like or plan-like attitude -- a state with the content like ‘I now plan to make a reflective inference and I won’t let any motivation in this process induce any action’ -- that prevents Pi from taking an action even when he has the motivation? Well, intuitively, in our case we see Pi simply sinks into the inference without a special plan about it.

---

4 Some people may point out that we don’t normally see a practically rational and normal agent who stays peaceful and conducts a pure reasoning when faced with a situation of life and death. I’d like to say that the observation doesn’t pose a real threat to my case. I totally agree that motivations and judgments are usually connected to each other empirically. But the problem for us is whether this connection is a conceptually necessary one. So what I need in my argument is a conceptually conceivable agent who is rationally competent. We can certainly assume that Pi is a rational and cool-headed person without so much chaotic irrational distractions. If the connection between moral judgments and motivations is really a conceptually necessary one, then the purely rational and cool-headed agent, Pi, is bound to have the appropriate motivation when he reaches (5). Actually, as for agents who are emotional and not cool-headed, they might have moral motivations even if no corresponding judgments are made, and that may actually mislead us in the debate.
Besides, even if we admit that Pi has such a plan, then according to Park (2016) we should expect that Pi undergoes a mental conflict when that plan obstructed his motivation associated with (5) from generating an action. However, what Pi actually does is just going on his reasoning peacefully at this point. Therefore, it seems to me more plausible to conclude that Pi makes the genuine moral judgment (5) while no corresponding moral motivation ever occurs.

Let’s now take the last criterion for a successful indifference argument into consideration: can externalists come up with a reasonable explanation for Pi’s case? To account for Pi’s case, externalists should be able to explain (i) why Pi has no motivation when making the judgment (5); (ii) why we intuitively expect him to have a corresponding motivation when he accepts (9); and (iii) why we should accept externalists’ explanation instead of internalists’.

As we know from the introduction, externalists claim that a moral judgment must come with some other distinctive desire-like mental states in order to cause an appropriate motivation. So externalists would be able to give a promising account for Pi’s case if (i’) Pi has no desire-like mental states to throw the fish into the sea when he judges (5); (ii’) it’s very likely for Pi to have a desire-like mental state to eat the fish when he judges (9); and (iii’) the account given by externalism is better than that of internalism.

Both (i’) and (ii’) are not hard to meet. When the cool-headed Pi judges (5) in the middle of his inference, we can imagine his attention all focuses on the reasoning that he is going through. It’s only after he has reached the final conclusion (9) that he’ll release his attention on the reasoning, allocating some of his mental resources to combine the conclusion (9) with his other desire-like states to generate a motivation.

As for (iii’), if there is really a conceptually necessary connection between the moral judgment (5) and the motivation to throw the fish back to the sea, then even if Pi is concentrated on the reasoning, he should still have the motivation to comply with his moral judgment and throw the fish back to the sea. But as we’ve analyzed above, this never really happens. On the other hand, according to externalism, during Pi’s reasoning when he has no mental resources allocating to relevant desires, he won’t be thus motivated. Externalists’ explanation sounds apparently more plausible.

Indeed, the intuition behind internalism is attractive: it’s so hard for us not to notice the widespread bound between moral judgments and motivations, especially in
cases when the judgments are derived from our memories, experiences, and moral intuitions, or where the judgments are used as reasons for or proclamations of an action. But if we claim from there that a conceptually necessary relation lies between making moral judgments and having motivations, we might be going too far.\(^5\) As my argument has shown, the connection between a moral judgment and a corresponding motivation, though widespread and strong, is more likely an empirical, external one than a necessary, internal one.

I’ve argued that ARMI satisfies all of the three criteria for a successful indifference argument, and I’ve addressed some potential objections along the way. Below I’ll consider how this argument could reply to internalists’ recent challenges and where its advantages lie compared with traditional indifference arguments.

### 3. Challenges and Responses

#### 3.1 Challenges deploying the avoiding strategy

How could ARMI respond to internalists’ recent challenges deploying the avoiding strategy? Let’s first consider versions of conditional internalism that limit themselves to psychologically normal agents, virtuous agents or practically rational agents (((Björnsson, G., et al. 2012)). The good news for ARMI is that Pi satisfies all of the above conditions, so such challenges won’t pose threats to ARMI.

But then, what if a version of internalism limits itself to moral judgments that close moral deliberations, ruling out all the moral judgments that are not conclusions of moral inferences? Well, since every moral judgment has the inferential property, it’s possible for any of them to be a premise in an inference. So if ARMI is plausible, then every moral judgment can at least possibly be made without generating a moral motivation. Even if moral judgments as conclusions of inferences appear to be accompanied with moral motivations, those judgments themselves are still possibly motivationally inert. Thus, there is no conceptually necessary bond between moral judgments and corresponding motivations, which means internalism is false. Indeed, as

\(^5\) I admit that there is an internal connection between moral judgments and motivations on the level of normativity – maybe we can say that the function or purpose of moral judgments is that they are most suitably used (sometimes with other distinctive desire-like attitudes) to generate motivations. What I don’t accept is that there are conceptually necessary connections between making moral judgments and having motivations on the level of individual cases.
we've discussed earlier, externalists can as well offer a concise and reasonable account for how the moral judgments as conclusions of inferences usually induce motivations, which makes this challenge even weaker.

Then could ARMI survive the challenge from Eggers’ potential version of internalism (Eggers 2015)? The answer may turn out optimistic, too. As we know, in the traditional cases concerning depression and extreme emotions, some obstructing elements – like the malfunction of the brain or the overwhelmingly strong emotions – might have prevented the potential of moral judgments from being realized (if there is really such a potential for moral judgments!), but in ARMI, no similar obstructing element has ever occurred. So how could the potential of moral judgments, if it makes any sense, be obstructed? And if the potential couldn’t be obstructed here, why is it not realized? The claim of ‘potential’ motivation seems empty in our case.

Another challenge is from Blackburn: traditional indifference arguments only undermine internalism when things are out of joint, but when things are not out of joint, the triumph belongs to internalism (Blackburn 1988). Fortunately, this is exactly where ARMI gains its extra advantage. Since my argument considers a normal and fully competent agent, things are not out of joint. Besides, since the inferential property covers all kinds of moral judgments, it’s natural to expand Pi’s case to form a systemic and unified objection to internalism – it’s possible for every moral judgment to be part of an inference, thus, it’s possible for every moral judgment to be motivationally impotent. In this way, we are not merely offering special conceivable counterexamples against internalism (like traditional indifference arguments try to do): we can also reject internalism in the main battlefield.

3.2 Challenges deploying the defusing strategy
One objection using the defusing strategy is to deny there is no motivation in the case. Internalists might say that the motivation occurs but is overridden, or is not acknowledged by the agent. They may also suggest that the motivation is suspended. In fact, these might be the most intuitive objections to ARMI. However, in Section 2, I have argued that, on the one side, there is no obvious source of motivational conflict or overriding in Pi’s case, and on the other side, externalists can come up with a very plausible explanation about why a motivation is generated only at the end of Pi’s inference. Those together should be sufficient to respond to internalists’ possible
objection here. Even if my argument there still couldn’t fully change internalists’ intuition. I hope it would at least make internalists reconsider their intuition.

Another objection would be that (5) is not a sincere moral judgment – either that Pi is not sincere in making it, or that Pi lacks sufficient capacity to fully understand it (Bromwich 2016). However, Pi’s rationality has guaranteed that he would be sincere when accepting (5), and it’s weird to think that a person who is fully competent of moral language would lack a sufficient understanding of the judgment that he acknowledges. If internalists must stick to the claim that having corresponding motivation is the necessary condition for one to understand a moral judgment, they would be offering a circle argument.⁶

The challenge from Swartze is also not a problem here. Swartzer contends that externalism cannot account for the depression case better by pointing out that people experiencing depression still have desires to fulfill their moral commitments (Swartzer 2015). However, in Pi’s case, if we ask Pi when he judges (5) whether he has the desire to throw the fish into the sea, we expect him to say ‘no’. This is different from the depression case where the victims claim to have real desires to conform to their moral judgments. We have also analyzed that Pi actually has no such desires when judging (5). Thus, such a challenge is not likely to pose a threat to ARMI.

Up to now, we’ve considered if ARMI can respond to each of the main challenges presented by internalists recently. We’ve seen that since ARMI resorts to a fully competent, normal and virtuous agent and argues from the universal fact that every moral judgment can appear in the middle of a series of moral inferences, it can survive internalists’ recent objections better than traditional indifference arguments do.

---

⁶ Bromwich insists this is not a bad circularity since internalists can ‘appeal to independent meta-ethical commitments to provide a non-question-begging response’ (Bromwich 2016, p. 8). Yet I think that’s quite untenable. The ‘independent meta-ethical commitments’ as Bromwich suggests, including cognitive/noncognitive semantics and the plausibility of Humean theory of motivation, are issues that face a lot of disagreements in metaethics. Even if we can take some of them as resources to defend Bromwich’s view, we can also take some of them to defend the negation of Bromwich’s view. Besides, motivational internalism gains its importance in metaethics mainly because of its role in the motivation argument, and in that argument, internalism and Humean theory of internalism are used as two premises leading to the conclusion of noncognitivism (Shafer-Landau 2003). How, then, could internalists take internalism as the reason for accepting noncognitivism on the one hand, and take noncognitivism as the reason for accepting internalism on the other?
4. Conclusion

In this paper I present an indifference argument from reflective moral inference (ARMI) by appealing to a fully competent, normal and virtuous agent. As I have argued, the new indifference argument will gain the upper hand when confronted with internalists’ potential and recent challenges, and it will be able to undermine internalism more systemically than traditional indifference arguments do.

References


Does a Physical Property Count as a Confounder?

XIN DONG

Jaegwon Kim’s Exclusion Problem is the problem of reconciling the causal efficacy of mental properties with the fact that they supervene on physical properties and the physical realm is causally closed. Kim maintains that the problem is insuperable, and so that non-reductive physicalism (NRP), according to which there is both mental irreducibility and mental causation, is untenable. Recently, however, some philosophers have argued that NRP theorists can appeal to James Woodward’s theory of causation and explanation — interventionism — to solve the Exclusion Problem.

Generally speaking, interventionism regards causation as a covariant relation, that is, X causes Y iff an intervention on X will bring about a change in Y. It has no trouble in arguing whether the causal relata is an event or a property, since it takes all of them as variables. As a result, all NRP need to do for preserving mental causation is to prove that an intervention on mental properties will bring about a change in other mental properties or the physical properties.

However, things do not go smoothly as expected since Woodward’s interventionism has an exclusive requirement on the process of intervention. In order to distinguish spurious causes from genuine causes, the latter of which is called a “confounder” because it confounds us to mistakenly attribute genuine causes’ causal power to the spurious ones, Woodward especially stresses that when we preform an intervention, the confounder must be hold fixed.

This requirement brings with it a serious challenge in the case of mental causation because, according to the principle of supervenience, physical properties cannot be hold fixed when mental properties in intervened. Therefore, if the physical supervenience base counts as a confounder, then we have good reason to doubt whether it is appropriate to discuss mental causation under the interventionist approach.

Therefore, if NRP insists on using interventionist approach to preserve mental causation and solve the exclusion problem, they have to, firstly, prove the compatibility of interventionism and supervenience. Aiming at this goal, Shapiro and Woodward have provided some arguments. However, in this paper, I want to illustrate that those arguments are still unsatisfactory and in order to ensure the compatibility, we may need to adjust the interventionist theory of causation.
In section 1, I will explain the specific character of Woodward’s interventionism. In section 2, I will present the incompatibility of interventionism and supervenience when applying the former on downward causation. In section 3, I will discuss two such would-be solutions (Woodward, 2014; Shapiro and Sober, 2010) and point out an assumption both of them make. In section 4, I argue that the assumption is highly questionable and I will also challenge the reasons they give for this common assumption. Therefore, I suggest that we’d better adjust interventionist theory in order to discuss downward causation.

1. The Character of Woodward’s Interventionism

The basic thought underlying Woodward’s theory is that if X causes Y, then we can change the value of Y by performing an intervention on the value of X. But this notion is not qualified enough if we don’t stipulate the intervention variable with more rules.

Consider the relation between the readings of a particular barometer and the occurrence of storm. If we use the atmospheric pressure, which is causally related to the storm, as an intervention variable, then, according to (M), this intervention on the readings of barometer will change the state of storm, which means the readings of barometer causes the occurrence of storm. Nobody will accept this absurd result.

This example reveals the situation of joint effects X and Y, of a common cause Z. Because of Z, X and Y mistakenly appear to be causally related. In this case, the common cause Z is a confounder, which make the spurious cause X look like a genuine cause.

With respect to this situation, Woodward added more conditions to rule out the disturbance of confounders and stipulate what counts as an appropriate intervention. Summarize as follows:

a) X should be caused and only caused by the intervention I. In Woodward’s word, I should cut off all the other causal paths toward X;

b) While causing X, I should not make a confounder change or be a confounder itself;

c) Z is a confounder for X iff Z causes Y through a different path from X (Z is off-path I-X-Y causal connection).
Combining the notions of causation and intervention variable, we can see that a confounder holds fixed by Woodward through two steps. Firstly, the confounder holds fixed when intervention variable I influences X. Secondly, besides intervention variable, there should not be other variables that make the confounder change. All in all, the confounder has no chance to disturb during the whole process of causal assessment.

2. Trouble in Solving Mental Causation

When NRP using interventionism to reserve mental causation, the main idea is that when M is changed by an intervention variable, M* and P* will change too, which entails that M is the cause of M* and P*. In this paper, I will not discuss this argument in detail, since what I concern is whether we can apply the interventionist approach to analyzing mental causation, not what will happen if we appropriately apply it.

As I stressed above, before analyzing a causal relation between X and Y, it is important to figure out two questions, that is, whether there is confounder Z in the variable set and whether we can hold the confounder fixed. Therefore, in order to analyze the relation between M and M* (or P*), we must start from clarifying these two questions.

Before clarification, I will briefly present the relational graph between physical and mental property, used by most of the discussion of mental causation, in order to make sure that we share the same presupposition.

As Figure 1 shows, all the NRP accept two principles:

Supervenience: mental properties supervenes on physical properties. (In this paper, I will also assume that multiple realization and supervenience are interchangable.)

Figure 1

The double line represents supervenience, the solid line represents causal relation and the dashed line represents the relation we need to assess.
The Causal Closure of Physical: every physical effect has a sufficient physical cause. Besides P-P*, whether the relations between M-M*, M-P* and P-M* are causal is begging the question. Comparing to P-M*, NRP concern more about M-M* and M-P*, hoping to save at least one of them. Since the situations of M-M* and M-P* are different, I will discuss whether it is appropriate to apply interventionist theory of causation in the model of M-M* and M-P* respectively.

Exam the model of M-M* causation first, using the two questions mentioned above: is there any confounder? Can we hold the confounder fixed?

![Diagram](image)

Figure 2

According to the definition of confounder in c, we should figure out whether P is a cause of M* before asking whether P is a confounder. Assume that m7 is multiply realized by p1 and p2 while m* is multiply realized by p1* and p2*. Then, we can have an intervention variable I change the value of P from p1 to p2 while other variable (which is M in this case) holds fixed. Since p2 cause p2* and p2* realizes m*, we find that the intervention does not bring about the change of M’s value. Therefore, P is not a cause of M* at all and cannot be a confounder of M. As a result, when analyzing the causal relation between M and M*, we needn’t worry about any disturbance from confounder.

Then, we should exam the model of M-P* causation, which is much more complicated than M-M* model.

---

7 In this paper, I will use capital number to stand for a variable and lowercase letter to stand for some specific value of a value.
As Figure 3 shows, P causes P* without any doubt. So, P is absolutely a potential confounder of M. If P is indeed a confounder, then applying interventionist approach to analyze downward causation will face serious challenge. In 2009, Michael Baumgartner raised this challenge by claiming that interventionism has two necessary conditions for X to be a cause of Y: manipulability and fixability. In his view, the supervenience principle violates these two conditions because it is impossible for P to be held fixed while performing intervention on M and we cannot find a qualified intervention I for the same reason.

Therefore, given supervenience, interventionists should pay more attention to the discussion of downward causation, since if P counts a confounder, then there is incompatibility between supervenience and interventionism.

3. Defending P for not being a confounder

Facing this challenge, interventionists try to prove that P is not a confounder with respect to M, just like the situation in M-M* causation. Let’s start with Lawrance Shapiro’s idea. He claimed that the covariance of M and P would not threat the causal power of M. He first used the Weismann’s experiment as an example and showed that the correct experiment method is to hold fixed the parental genotype while testing the causal power of parental phenotype toward offspring genotype, not to fix the micro-supervenience base of that phenotype. He warned us that we should be very careful when deploying “holding-fixed arguments”. And he concluded that:

Investigating whether X causes Y involves figuring out whether wiggling X while holding fixed whatever common causes there may be of X and Y will be associated with a change in Y. It is not relevant, or even coherent, to ask what will happen if
one wiggles X while holding fixed the micro-supervenience base of X. (Shapiro and Sober 2007, p.6)

The reason why trying to hold fixed the micro-supervenience base is a wrong manipulation is that causation is a diachronic relation. Therefore, a real confounder $Z$ should be a common cause that occurs at different time from both $X$ and $Y$, not just from $Y$. According to this stipulation, the micro-supervenience base cannot be a confounder since the (mereological) supervenience relationship is synchronic. Later, Shapiro stressed again that “the right test to perform will hold fixed, not the supervenience base of $M$, but the common cause that $M$ and $P^*$ share”. (Shapiro 2010, p.600)

Similarly, Woodward argued against Baumgartner’s challenge in his newest paper. (Woodward 2014) Basically, he has the same assumption with Shapiro that $X$ and the confounder $Z$ should have causal relation, that is, a confounder $Z$ should be a common cause of $X$ and $Y$. Unlike Shapiro, Woodward gives out more reasons why we should not take the supervenience base $P$ as a confounder of $M$.

First of all, it is possible to set two variables in a set $V$ to different value via two independent interventions if these two variables have causal relation while it is impossible to do so if they have non-causal relation, such as definitional, logical, mathematical, mereological or supervenience relations, etc.. Therefore, if non-causal relation occurs in a set of variable $V$, then the notion of intervention variable and the requirement of fixability should be adjusted in order to fit the covariance of non-causal relation.

Then, Woodward further illustrates why a variable in non-causal relation is not a confounder. He takes conceptual dependence as an example. Suppose that heart disease ($D$) is causally influenced by two variables: high density cholesterol (HDC) and low density cholesterol (LDC). The former lowers the probability of disease while the latter raises it. Besides, there is another variable: total cholesterol (TD), which is defined to be the sum of HDC and LDC and can causally influences $D$ as well. The relationship between those four variables is shown in Figure 4 as below:
Figure 4. The double-tailed arrow represents conceptual dependence while the single-tailed arrow represents causal relation.

If we don’t distinguish causal and non-causal relation, then we should hold fixed both LDC and TC when testing the causal power of HDC toward D. However, as Woodward claims, LDC should be held fixed while TC should not. To summarize, the change of TC has three disparate features while LDC is held fixed and HDC is changed:

i) The variation of TC is tantamount to the variation of HDC;
ii) The influence TC brings to D is just the same as the influence HDC brings to D.
iii) The causal arrow from TC to D is not indispensable. Having this arrow or not are “simply different representations of the same causal structure” (Woodward 2014, p.40).

Therefore, although TC and D have causal relation, TC doesn’t cause D through a different path from HDC. That is, TC is not off-path the connection between HDC and D.

As Woodward states, the case of supervenience is just the same as conceptual dependence. So, “it is inappropriate to control for supervenience bases in assessing the causal efficacy of supervening properties… On the interventionist conception of cause, when non-confounded manipulation of your mental states is associated with changes in cognition and behavior, this is all that is required for mental causation.” (Woodward 2014, pp.19-21)

To conclude, both Shapiro and Woodward think that the supervenience base of M need not be held fixed. In my point of view, they share the same assumption that only common cause can count as a confounder although they have different reasons. In the next section, I will doubt this assumption and demonstrate that we cannot easily rule supervenience base out as a confounder.

4. The failure of the defense

Both Shapiro and Woodward share the same assumption that P is not off-path the I-M-P* connection since P is not a common cause of M and P*. However, whether Z is a common cause should not be a criterion. Indeed, most of the time, a confounder is the
common cause of X and Y, but it does not mean that only common cause can be a confounder. Provided Z causes Y through a different causal path from X, Z counts as a confounder, no matter whether the relation of X and Z is causal or non-causal.

Since not-a-common-cause cannot help supervenience base P to avoid the suspicion of being a confounder, we should discuss more about the detailed reasons that Shapiro and Woodward give, in order to exam whether P is a confounder, that is, whether P is off-path the I-M-P* connection.

As mentioned before, Shapiro states that a confounder Z should occur at a different time from both X and Y. I agree that the time of Z’s occurrence should be different from Y in order for it to be a cause of Y. But it is not clear why the time of Z’s occurrence must be different from X’s. Occurring simultaneously is not conflicted with Z’s causing Y through a distinct path from X. It seems that the burden falls on Shapiro’s shoulder to prove that two variables that in a synchronic relation cannot have different causal paths toward the same variable.

Consider Woodward’s defense. The main argument Woodward gives is that P and M are in non-causal relation. And he uses conceptual dependence on behalf of this kind of relation to explain why a non-causal variable is not off-path with three reasons. I totally agree that conceptually dependence satisfies these reasons and these reasons entail that conceptually depending variable is not off-path.

However, I doubt whether supervenience is the same as conceptual dependence and whether it can also satisfies those three reasons as well. These are two issues because many relationships that are different from conceptual dependence but still satisfy these three reasons and, therefore, is not off-path.

First, I want to enumerate several differences between supervenience and conceptual dependence in the chart below.

<table>
<thead>
<tr>
<th></th>
<th>Multiple realization (NRP’s sense)</th>
<th>Conceptual dependence</th>
</tr>
</thead>
<tbody>
<tr>
<td>One-one correspondence</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Changing quantity</td>
<td>Not the same</td>
<td>The same</td>
</tr>
<tr>
<td>Independent fixability</td>
<td>Half no, half yes</td>
<td>no</td>
</tr>
</tbody>
</table>

The first line need not be explained too much because the conclusion is derived from the definition of these two relations. According to multiple realization, the realized
variable can be realized by many realizing base. So, it is more like a one-to-many correspondence. By contrast, concept dependence is one-to-one. It is not to say that a variable can conceptually depend on only one variable. Like the example Woodward gives, TC conceptually depends on two variables, HDC and LDC. But when LDC (or HDC) is given, the value of TC and the value of HDC (or LDC) is one-one correspondence.

The second line can be, in fact, inferred from the second one. Whenever the depended variable changes by certain quantity, the depending variable will change by the same quantity and vice versa. Compared to this, if the realized variable change from m to m*, we are not sure whether the realizing variable change from p1 to p1*, or p1 to p2*, or p2 to p1*, or p2 to p2*. Since one change in realized variable may accompany with many possible changes in realizing variable, the quantities of which are different, we can safely conclude that the changing quantity between these two variables are different.

With respect to the third line, I agree with him that conceptual dependence relation cannot satisfy the principle of independent fixability completely, since once we fix the depended variable, there is no other independent intervention variable that can set the depending variable to any of its individually possible values, and vice versa. However, multiple realization cannot satisfy this principle as perfect as conceptual dependence because when the realized variable, like m, is held fixed, it seems that there can be another independent intervention variable that can set the realizing variable to several of its specifically possible values, like p1 and p2.

These three differences have shown that the situation of conceptual dependence and that of supervenience are not the same. Therefore, the conclusion, which is suitable for the former cannot be applied to the latter. Then, let’s further analyze whether supervenience (multiple realization) can satisfies those three reasons that Woodward gives. Firstly, it is easy to say that the variation of M is not tantamount to the variation of P unless someone just bite the bullet and say that the changes from p1 (p2) to p1* (p2*) and from p1 (p2) to p2* (p1*) are equal to each other, which is really counterintuitive.

Secondly, we have plenty of examples to show that the realized variable and the realizing variable can have different influence upon the same variable. Take Yablo’s famous pigeon for example. It is easy to conceive the following two situations: seeing
red make the pigeon peck while seeing scarlet cannot counts as the cause of pecking, or, seeing scarlet make the pigeon peck, not seeing red.

Thirdly, the causal arrow from the realized variable to an effect is indispensable. Since the causal influences from the realized and realizing variable are different, the causal arrows from both of the variable show something exclusively. Take the pigeon story for example again. The causal arrow from seeing red to pecking and that from seeing scarlet are not just different representation of the same causal structure. We cannot deduce the former from the latter, and vice versa.

All in all, supervenience cannot fit those three reasons as perfectly as conceptual dependence do. Furthermore, after analyzing all those differences between these two non-causal relations, it seems that P causes P* through a distinct causal path, not sharing the causal path with M-P*. Therefore, we may conclude that P is off-path the I-M-P* connection, which make it a qualified confounder. And according to interventionism’s requirement, P should be held fixed while performing intervention on M, which is violated by the principle of supervenience. Therefore, interventionism and supervenience is incompatible in the case of downward causation.

5. Conclusion

Through this paper, I have shown that interventionist theory of causation stresses that a confounder must be held fixed when we assess causal relation. However, in downward causation, every mental property has a supervenience base, the physical property, and the latter has its own causal power toward other physical properties. So, we have to demonstrate whether physical property counts as confounder. That is, whether physical property causes other physical properties through distinct causal path.

Shapiro and Woodward point out that since supervinience is not a causal relation, P is not a common cause of M and P* and therefore, is not a confounder at all. Furthermore, Woodward uses conceptual dependence, on behalf of supervenience, as an example to illustrate why it does not count as confounder by giving out three reasons.

However, I argue that whether it is a common cause cannot determine whether it is a confounder. Through clarifying the difference between supervenience and conceptual dependence, I illustrate that the three reasons given by Woodward is not suitable for
supervenience and P is off-path the connection I-M-P*. Therefore, I claim that the physical counts as a confounder with respect to the mental when assessing downward causation.

Then, we can conclude that the principle of supervenience violates the requirement of fixation because we cannot hold fixed P when performing intervention on M. And this incompatibility shows that we should adjust interventionist theory of causation, using other approach to rule out the disturbance from confounders. With this renewed version of interventionism, then, we can investigate whether mental property causes physical property.

References


How Could Agent-Causal Theories Survive?

ZHAO YI

Confronted with severe challenges posed by the Libertarian Dilemma, which suggests free will should be, in order to be neither determined by events nor happened merely by chance, not only incompatible with determinism but also compatible with indeterminism, libertarians find agent-causal theory a possible way out. Similar to earlier versions such as Kantian “noumenal self” and Cartesian “soul”, agent-causal theories require stronger metaphysical stipulations compared to their event-causal counterparts. Thus criticism often amounts to accusations of the obscurity and mysteriousness of the idea of agent causation. But in the past decades, contemporary agent-causal theories have been developing with increasingly detailed accounts of agent causation (chiefly by Richard Taylor, Roderick Chisholm, Randolph Clarke, Timothy O’Connor). In the meantime, more specific objections are correspondingly occasioned (chiefly by C. D. Broad, Peter van Inwagen, Alfred Mele). These more recent accounts and objections will be the main resource of this essay’s discussion – how could agent-causal theories survive?

1. Theoretical Obligation and Ontological Commitments

In the mode of event-causation which functions directly from circumstances to effects, every event is caused and therefore determined by some prior events. For the incompatibilist libertarians, it is hard to imagine free will (in a sense which makes us morally responsible for our decisions and actions) in a world where every decisions and actions as events are caused and determined by some prior events rather than ourselves, meanwhile, it is even more dangerous to consider human decisions and actions as not caused at all. Neither of these states could guarantee our free will in the sense that we are morally responsible for what we did. If human choices and actions cannot be caused by any internal or external events but still be caused anyway, it seems that there is only one solution left – it is caused at least in some degree by us – the agents.

Therefore, agent-causation, according to the chief agency theorists, are basically defined through event-causation, as “another species of the causal genus” (O’Connor 2000, 72) contrasted with event-causation, or more precisely, it is “an ‘ontologically primitive’
notion of causation by an agent or substance that is not reducible to ordinary modes of event-causation” (O’Connor 2011, 309). But it seems insufficient to support the assertion of agency barely by this “other than event-causation-ness”. Without a detailed ontological foundation, recent agency theories cannot survive the wide-spread distrust in what they labeled as agent-causation. Generally speaking, the posited definition of agent-causation requires at least a tri-level ontological commitment. This essay bases on the account of Timothy O’Connor – the most important contemporary agent-causationist, as the ontological foundation of agent-causal theories. 

First, agent as substance should endure through time. A temporally extended object appears to be incompetent to play the role of an agent that is persistent and morally responsible through one’s whole life. Events occur and exist at a particular point of time, while “agent-causal theories require that we think of objects (and agents specifically) as things that endure through time, such that they are wholly present at each moment of their existence, without being identical to the total state they are in at that moment.” (O’Connor 2011, 312) 

Secondly, agent should be compositionally irreducible. If agents are objects constituted by certain constituents, the decisions and actions they caused would be traced back to those constituents as prior events. Thus, agents need to be irreducible substances—entities that are more than the sum of the constituents of their bodies. But how could human beings possibly be physically composite while compositionally irreducible? O’Connor proposes that “agent causal power and its allied properties are ontologically emergent, although still powers and properties of the biological organism” (2011, 313), through which stipulates agents’ compositional irreducibility ontologically. 

Thirdly, agent-causation requires an irreducible concept of causation. The core and primitive element of the irreducible thesis is the “producing” or “bringing about” or “derivativeness” of an effect from its cause. By contrast, the reductive Humean analysis of causation reduces the relation of causation into noncausal facts, such as certain patterns of similarity among event types, without the “derivativeness” of the effect. The reduction of causation also eliminates the controlling power of agents as the cause in a chain of causality. Clearly, agent-causation should be understood more than mere patterns of similarity or any such terms.
We can see that either the specific meaning or the ontological foundation of agent-causation should be developed in great carefulness on the foundation of these theoretical obligations. However, before agent-causal theories strives to meet the demand of those theoretical obligations from the Libertarian Dilemma to make free will be able to be morally responsible for our decisions and actions, there is a factual premise that should be first taken into consideration. Commonsensically, we think that an agent (in the sense that she is morally responsible) should be act on the basis of her own reasons, which amount to the states of agent-involving events. If the agent’s decisions and subsequent actions cannot be caused by, the states of having a variety of reasons, as events, the problem arises of how to explain the relationship between agent and reasons and this agent-involving reasons-agent-effects mode. Hence, in the next two parts, we will examine agent-causal theories with their most decisive challenges from agent-reason relationship on the one hand, and the accomplishment of their libertarian theoretical needs on the other.

2. Agent-Reasons Relationship and the “Datable” Challenge

Even if all of those ontological commitments are accepted by the opponents that the notion of agent-causation is not totally intelligible, there still remains doubts upon the possibility of the reasons-agent-effects cooperation such that it is impossible for an agent as a substance to cause anything. The best-known objection for agent-causation concerns the timing of causal effects is raised, in the middle of this century, by C. D. Broad, which is taken by many to be decisive:

I see no prima facie objection to there being events that are not completely determined. But, in so far as an event is determined, an essential factor in its total cause must be other events. How can an event possibly be determined to happen at a certain date if its total cause contained no factor to which the notion of date has any application? And how can the notion of date have any application to anything that is not an event? (1952, 215)

Broad points out that while an event has to be caused at least by a factor to which the notion of date can be applied, agent as substance is not supposed to contain any this kind of factor; and if agent as substance is understood as containing this factor to which the notion of date can be applied, it is not supposed to be an agent any longer for the timing factor is exclusive for events not agents. By introducing the timing factor, Broad
demonstrates the apparent impossibility of the reasons-agent-effects pattern. Truly, in the ordinary sense, agents are not “datable” objects. Nevertheless, according to contemporary agent-causationists, this type of analysis is senseless since only agent can hold itself responsible for a certain effect and those facts that are true of an agent at one time do not hold of her at another. O’Connor illustrates this point by an example:

Consider, for example, my deliberation concerning whether to continue working on this chapter or to stop and do something else. After considering the matter, I formed the intention at time $t$ to continue working. We may suppose that at $t$ I had the power to choose to continue working or to choose to stop, where this is a power to cause either of these mental occurrence...truthfully that Tim at time $t$ causally determined his own choice to continue working. We needn’t, to make sense of this, analyze it as the claim that a “datable entity”, Tim-at-$t$, was the occurrent cause of the decision to continue working. (2000, 75)

Here O’Connor denies that agents need to be “datable entity” to cause an effect with timing factor while the timing factor can be inherited from the reasons involved in the process of agent’s free choosing, which means that “the agent’s capacity to cause action-triggering events is causally structured by the agent’s internal state, involving the having of reasons and other factors, before and up to the time of the action” (2011, 319). But this explanation leads to the demand of a further description of this reasons-agent-effects pattern. How is agent influenced by reasons and how are those effects finally caused? O’Connor proposed an account that employs Fred Dretskes’s distinction between structural and triggering causes. To illustrate the distinction, the structuring cause of an explosion is the whole process by which the bomb is constructed, while its triggering cause is the lighting of the fuse. In O’Connor’s point of view, reasons are structuring causes of a decision which structure the propensities of the agent-causal power whose exercise produces it, while the agent plays the role of the triggering cause in her exercise of this structured power. The agent, casually structured by reasons, exercise her triggering power as the direct cause, generate a range of intentioned effects as events. As O’Connor himself puts it:
While nothing produces an instance of agent-causation, the possible occurrence of this event has a continuously evolving, objective likelihood. Expressed differently, agent-causal power is a structured propensity towards a class of effects (the formings of executive intentions), such that at any given time, for each causally possible, specific agent-causal event-type, there is a definite objective probability of its occurrence within the range (0,1), and this probability varies continuously as the agent is impacted by internal and external influences... The effect of influencing events is exhausted by their alteration of the relative likelihood of the outcome, which they accomplish by affecting the propensities of the agent-causal capacity itself. (2009, 15)

If, as O’Connor illustrates, reasons, by changing propensities of the agent-causal power, structure the agent’s triggering power to cause certain effects, then is this structured agent’s triggering power strong enough to hold the agent per se morally responsible? For agent-causationist as O’Connor, surely it is:

We insist upon the distinction between reasons structuring one’s agent-causal power in the sense of conferring objective tendencies towards particular actions, and reasons activating that power by producing one’s causing a specific intention. On the view I have described, nothing other than the agent himself activates the causal power in this way. (2009, 38)

In O’Connor’s view, whatever the objective tendencies or propensities towards particular actions are, it is the agent who has the final say on whether to activate those particular actions or not. To put it in another word, even if I have an objective probability of 0.999 to cause the intention to attend class, I may still fail to attend it. I, as the agent, always have my power to resist this strong propensity as well as act upon it.

This claim is intuitively right, because we do feel the same way when we are making decisions in our daily life: whatever the choices are, and however the reasons work in my mind, it is me who finally makes the decision and directly causes the effects. In order to maintain agent’s predominant triggering power as the cause, agent-causationists need to, more or less, appeal to this epistemological evidence from direct personal experience. As Thomas Reid, who is generally regarded as the father of modern agent-
causal theories, says: the conception of an efficient cause may very probably be derived from the experience we have had ... of our own power to produce certain effects”. Although the “datable” problem challenges the possibility and plausibility of agent-causal theories’ agent-reasons relationship, agent-causationist do not fail to produce a sound explanation of the reasons-agent-effects working pattern. But as long as reasons involve in the process of decision-making in tandem with agent, it is hard to determine which is supposed to be the cause that takes the moral responsibility unless invoking another epistemological assumption directly from our own experience.

3. Luck Problems

Challenges like the “datable” problem think that agent-causal theories are incoherent because of the intrinsic conceptual weakness, but many thinks that even though agent-causal theories can be intrinsically coherent with all their ontological and epistemological stipulations, they cannot be compatible with the outcome of indeterminism and therefore accomplish their theoretical commitments, so they would remain useless anyway. Opponents raised an array of challenges of this type. This essay would collectively refer to them as Luck Problem.

3.1 The “rollback” problem

Peter van Inwagen’s principal argument for this contention has come to be known as the “rollback” argument. He asks us to imagine an agent, Alice, who chooses to tell the truth when be able to choose either lying or telling the truth. And as van Inwagen proposes:

let us assume that free will is incompatible with determinism, and that Alice’s telling the truth, being a free act, was therefore undetermined. Now suppose that immediately after Alice told the truth, God caused the universe to revert to precisely its state one minute before Alice told the truth (let us call the first moment the universe was in this state ‘t1’ and the second moment the universe was in this state ‘t2’), and then let things "go forward again." What would have happened the second time? What would have happened after t2? Would she have lied or would she have told the truth? (2000, 14)

Consistent with the posited agent-causal theories – Alice’s decision is undetermined, so her second choice would also be undetermined. Now suppose that God were to enable
1000 times of this kind of “rollback” to exactly the state it was in at t1. Still, at each time, Alice’s decision was undetermined, such that we cannot tell whether she would lie or tell the truth. But as the number of “rollback” increases, “we observers shall almost certainly observe the ratio of the outcome ‘truth’ to the outcome ‘lie’ settling down to, converging on, some value.” (2000, 14) So as van Inwagen continues to ask: “Is it not true that as we watch the number of replays increase, we shall become convinced that what will happen in the next replay is a matter of chance?” (2000, 14)

This “rollback” thought experiment depicts a similar picture as flipping a coin. The outcome of every single flip is undetermined while the ratio of the outcome will settle down to a certain value. What van Inwagen suggests here is that the agent’s decision-making of agent-causationists is identical to flipping a coin, merely “a matter of chance”, such that this would not count as a free act.

This argument cannot give a fatal blow to agent-causationists, not only because we do not have a God to roll back time either in actual case or from the known physical laws, but the argument is built on the premises that agent-causationists would not admit. As Meghan Griffith contends, while given the event-causal conception this kind of reasoning is plausible, on the agent-causal view it is not. (2005, 263) For even if the ratio of the outcome would eventually settle down to a certain value, Alice, as the agent still have the power to determine whichever the decision to be made.

### 3.2 The “strengthened” luck problem

Mele advances a strengthened version of the luck objection. Suppose that there is some other possible world W*, where all the causal antecedents of D (agent A causes D at t) in the actual world W are present. But while D occurs in W, it fails to do so in W*. Thus, the fact that D does come about would seem to be a matter of luck. The argument seems no different from the rollback problem and the infinite regress challenge, agent-causationists could resort to agent’s direct triggering of the choice again, whether the outcome of the choice-making process seems to be by luck or not. But Mele’s strengthened luck objection takes a further step:

... the question why an agent exercised his agent-causal power at t in deciding to A rather than exercising it at t in any of the alternative ways he does in other possible worlds with the same past and laws of nature is, in principle, unanswerable ...
because there is no fact or truth to be reported in a correct answer ... and his exercising it at t in so deciding has an effect on how his life goes, I count that as luck for the agent. (2006, 70)

Why would the agent choose D in W but not any other alternative ways he would choose in W* given exactly the same past and laws of nature? It is “unanswerable”, even agent-causationists would agree that it is because no reason here could be given (all the reasons could be given are exactly the same in the two worlds). Further, agent-causationists would hold that it is totally “up to the agent”, so at this stage, no reason could explain the occurring of D or not-D. But Mele could equally raise a construal in an utterly different way: no reason could be given just means there is no reason. Even if D is brought about by agent’s exercise of her agent-causal power, clearly it is not out of reason, therefore it is by luck. Or, to say that agent’s exercise of her agent-causal power is no other than to say that the decision is occurred by luck, “because there is no fact or truth to be reported” in this case.

We can see that, Mele’s notion of luck is not necessarily conflict with moral responsibility or free will, which means that actually he does not intend to deny free will in the sense that it is morally responsible by saying that the decision is made by luck. As O’Connor comments, Mele “does not press his luck objection as a deep skeptical worry about indeterministic freedom (as van Inwagen does with his rollback argument). Instead, he wields it to neutralize the agent causationist’s objection to causal indeterminism ... His aim is to show that the agent causationist gains nothing by positing a distinctively agential capacity of control” (2011, 325).

In reply, agent-causationists like Clarke (2005) thinks that the agential capacity of control posited by agent-causation is imperative, for it proffers a stronger variety of control compared with that is available on Mele’s accounts. And the notion of luck, in Mele’s sense, cannot bring about the moral responsibility and free will that libertarians are seeking.

This essay will not discuss why the causal indeterminist advocated by Mele cannot get libertarians the free will they want, nonetheless, the problem, again, gets back to, not what the truth is, but what the libertarians need. We get back to the most common criticism for agent-causation theories, that, as Gary Watson comments, “agent-
causation’ simply labels, not illuminates, what the libertarian needs”. (2003, 10) Then, even if agent-causal theories are valid within the libertarian theoretical stipulations, it is in question that whether the libertarians’ theoretical needs are reasonable, and whether it is valid to accept all the libertarians’ stipulations to get what the libertarians’ needs. After all these discussions and arguments back and forth, we can see that, the plausibility of agent-causal theories is, at least partly based on, the theoretical demands that at the very first place give rise to them.

4. A New Perspective

The reason why agent-causal theories would evoke such heavy criticism largely lies in people’s bare imagination of “another species of the causal genus” rather than event-causation. Although we all have our own experience of exercising agent-causal power, event-causation appears to be too universal and all-encompassing to be challenged. No matter it is like “datable” problem which doubts upon the possibility of the cooperation of event and agent, or like the luck problem series which suggests the causal power of agent parted from events would definitely fall into randomness, essentially all the objections just comes from the impotency of imagining agent-causation in the face of its powerful event-causation counterpart. Truly, the commonsensical event-causation is absolutely the law of nature, whilst our recognition of agent-causation is merely “unexamined” mental feelings.

As a result, many, even agent-causationists themselves, when they need to analyze how agent A works to cause decision D, incline to hold that A is structured by reason R to cause D. In such analysis, A as agent cannot exercise its causal power with any event as causes, which makes A’s decision appear to be senseless and irrational, while R as event seems to have a stronger sense of causal power, thus it is seemingly plausible to think that R could independently cause D without A’s causal power. So at this stage, A’s posited agent-causal power seems to be too weak compared with that of R. For example, agent Ming, structured by reason “students should attend class”, directly causes event “Ming attends class”. Supposing reason “students should attend class” is absolutely undoubted in Ming’s mind, then it seems that it is this reason that causes the event to a greater extent rather than the agent. And reason “students should attend class” could be caused by other reasons, like the precedent reason “students should be hard-working.
And all of these reasons could be derived from Ming’s upbringing and growing environment, but neither of these could be determined by Ming himself as an agent.

This kind of reasoning appears to be very natural in arousing suspicion of agent-causal power. But when it is dealt with more carefully, another piece of important detail emerges. External factors do influence the agent, but this kind of influence occurs only after agent’s choosing to be influenced by exercising its causal power. Similarly, reasons in agent’s mind could be the cause of a certain decision, but whether they are to be used in a particular decision depends on whether agent choose to accept them as reasons in the first place. Like Ming’s being educated of “students should work hard” cannot directly lead to the reason “students should work hard” in Ming’s mind, only after the acceptance by Ming’s agent-causal power could the external factor truly effect the agent with the reason “students should work hard”; and the reason “students should work hard” must also be accepted by the agent’s causal power to really come into play. It is clear that, the acceptance of agent is the first condition for any circumstances to come into the agent’s mind as a reason. In order to decide whether to accept a certain reason or not, agent might need another reason for this reason. But this reason’s reason is still established after the “accept-or-not” dual choice. Therefore, however reasons or even external factors are required for a particular decision, agent is the determining cause of each stage. In addition, since agent does not need to be structured by any other conditions to make decisions but purely make the “accept-or-not” dual choice. So the agent could play the absolutely determining role as a cause without appearing to be an empty conception. Hence, this essay thinks a complete expression of the decision-making process involving agent-causation is:

\[
D \iff A(Rn) \iff A(Rn \lor \neg Rn) \iff A(Rn-1) \iff A(Rn-1 \lor \neg Rn-1) \iff \ldots A(R1) \iff A(R1 \lor \neg R1) \iff A(I) \iff A(I \lor \neg I)
\]

In this expression, \( \iff \) stands for causal relation with the arrow pointing from the cause to the result; \( A(Rn) \) stands for agent A’s choosing \( Rn \) as reason; \( A(D \lor \neg D) \) represents that agent A’s choosing from \( Rn \) or not-\( Rn \), \( A(I \lor \neg I) \) represents that agent A’s choosing from to be affected by external factor’s influence I or not-I. From this expression, we can see that, both reasons inside and factors outside have to go through agent’s dual choice in order to obtain some causal power in the decision-making process. Therefore, the
production of a rational decision is the outcome of a series of dual choices; agent-causal power avoids to be determined by any events or reasons as well as being totally empty, as agent’s causal power purely lies in the “accept-or-not” dual choice.

Nevertheless, if the reason of a dual decision comes from another dual decision, in this manner, there has to be a first reason which is derived from no other reason, like the end of the expression above -- \( A(I) \iff A(I \lor \neg I) \). And to illustrate by the example, why does Ming choose to accept the “hard-working” doctrine from the dual choice in the very first place rather than simply rejecting it? Is it true that, this initial dual choice, fall into randomness eventually? This essay holds that, agent’s initial dual choice could be derived from its innate inclination, or even determined by its biological conditions such as gene. However, when an agent is making a decision under the influence of its innate biological inclination, the agent finds it acceptable and satisfactory, or even feels that this type of innate inclination is exactly one part of agent per se. So the innate inclination could not deny agent’s free will only because it is not created by agent per se.

Likewise, agent-causation theories could admit that reason R determines decision D, which would not defect agent’s free will for agent’s free will lies in reason R itself. Thus, agent A in world W would not choose decision D out of no reason compared with decision not-D being chosen in world \( W^* \) given exactly the same past and laws of nature; but to choose on the basis of reasons, for reasons are not things that determine agent, but merely one part of the agent-causal power that was accepted by the agent in the first place.

### 5. Conclusion

Agent-causation theories can generally ensure a morally responsible free will on the foundation of its stipulations after a barrage of criticism. But, it is challenged that, even if agent-causation theories survive all the objections and are theoretically coherent, whether it is the real answer for free will question. From a more meticulous scrutiny on the occurrence of reasons, this essay finds that reasons per se could be part of the agent-causal power, such that frees agent from being determined or random and therefore ensures free will required.
Hence, through an increasingly elaborated explanation of agent-causal theories’ stipulations, the agent-causal theory proves itself to be a consistent system to ensure the free will required. But closer examinations of all those foundations are needed, otherwise the magnificent building of agent-causation might also be a mirage.

References


SITUATIONISM AND MORAL RESPONSIBILITY

ZHANG KE

Introduction

Situationism has been widely discussed over the past few decades, and it has already had its implications on moral theories and our moral common sense. Its impact is both theoretical and practical, ranging from virtue theories and moral responsibility to moral practice in real life, such as moral evaluation regarding wrongdoings in wartime and mistreatment towards minorities due to implicit bias. Among the earlier discussions, the series of social psychology experiments were construed as revealing how problematic it is to hold a characterological interpretation and overlook situational factors in explaining human behaviors (Nisbett & Ross, 1991; Harman, 1999; Doris, 2002). Later on, the overlooked yet interactive relation between the normative dispositions internal to the agents and the situational factors external to the agents triggered off the skepticism towards moral responsibility, indicating the normative competence, that is, the capacities to respond to moral considerations required by moral responsibility, might be compromised under the influence of situational factors (Nelkin, 2005; Brink, 2013; Vargas, 2013).

The main challenge from situationism to moral responsibility concerns this paper is that situationism compromise either the possession of normative competence or the exercise of normative competence. Among different responses, based on the distinction between possession and exercise, I will discuss two replies in this paper, and try to show that neither of them serves as sound reply unless we make certain clarifications and lose certain concepts of the morally responsible agent.

In section 1, I will illustrate the challenge from situationism in detail. In section 2, I will introduce the argument from possession (Doris, 2002; Brink, 2013), which is an approach to save the possession of normative competence, and in turn, the concept of the morally responsible agent, by compromising the exercise of normative competence.

Note that there have been discussions on whether character assessment presupposes moral responsibility evaluation. I do not intend to discuss this topic here, though I tend to be sympathetic to the idea that moral responsibility evaluation can be made independent of character assessment. See the discussions in Doris (2002), and Brink (2013).
After discussing its pros and cons, section 3 will be focusing on another argument from exercise. In section 4 I will come to a conclusion that either we admit that the concept of the morally responsible agent is redundant in the practical sense, and embrace one that is based on the exercise of normative competence, or we do not have a concept of the responsible agent at all, and for the actors in variety of human behaviors, the moral status as responsible agents is a sham.

I. Challenge from Situationism

Situationism started from a series of works in social psychology, which test the interactive relation between situations and the supposedly patterned behaviors among people. Although the interpretations vary in terms of what situationism’s implications are, there are agreements to some degree on the overlooked role situational factors may play in shaping human behaviors, especially the insignificant ones. This kind of overlook is also called “the fundamental attribution error” by Gilbert Harman along with other social psychologists (Ross & Nisbett, 1991; Harman, 1999). In addition to the moral evaluations that people make on what brings out certain behaviors, that is, people tend to attribute behavioral outcomes to the normative structure internal to the agents, not giving situation its due credit in explaining human behaviors. This error also appears to the agents in action, indicating that there are situational factors that, on the one hand, constitute morally salient information, and on the other hand, are insubstantial in the sense that they are beyond the agents’ awareness and the unawareness is somehow in a causal relation with the behavioral outcomes. The latter interpretation of the error casts doubts on whether insubstantial situational factors have any causal power regarding the initiation of certain behaviors, and if they do, does our moral standing as responsible agents degrade or even dissolve for that reason?

It has been suggested that the empirical findings brought about by situationism poses threats to moral responsibility, and especially to the view that moral responsibility

---

9 Dana Nelkin has a different interpretation concerning the implications of the fundamental attribution error and those of situationism. She suggests that former implies that people in general attribute actions to the wrong things. For example, when they use characterological explanations when it is situational factors cause an actions. While situationism implies that people misidentify the causal role certain situational factors play in bring out actions (Nelkin, 2005).
requires the agents' sensitivity to good reasons (Nelkin, 2005; Talbert, 2009; Brink, 2013; Vargas, 2013; McKenna & Warmke, forthcoming). In the presence of insignificant situational factors, such as the ones that could give people mood swings (Isen & Levin, 1972), or mere instructions from a seemingly authoritative figure (Milgram, 1969), or even situational stimuli that reflects the race or gender, people tend to act inconsistently regarding what they say or think they believe. One interpretation of this phenomenon, which I adhere to for the purpose of this paper, is that the actors in question are not acting for morally salient reasons. And if we assume that being sensitive to morally salient reasons, or to be more specific, the capacities to 1) recognize morally salient reasons, and 2) transfer those reasons into actions\textsuperscript{10}, are implicated in the concept of normative competence, and in turn, the possession and exercise of normative competence are crucial to the concept of the morally responsible agent\textsuperscript{11}, then situationism challenges moral responsibility by showing how constantly we are a less responsible agent.

Here I consider the relation between normative competence and moral responsibility, especially the concept of the morally responsible agent, by focusing on a pair of concepts, i.e., “possessing normative competence” and “exercising normative competence.” Possession is necessary for moral responsibility since the moral standing as a responsible agent is issued from the status of possessing certain responsibility-relevant capacities. That is, if it is reasonable to see one as the possessor of normative competence, correspondingly, it is also reasonable to see her as a morally responsible agent who is responsible for what she does. Exercise also plays a part in understanding moral responsibility since it is correlated with both possession and the responsible act. That is, it is reasonable to assume that the possessor of normative competence regularly exercises her normative competence. In a nutshell, possession indicates a state of inactivity, exercise indicates an active state of using what one possesses. This makes

\textsuperscript{10} Similar idea of pairing responsibility-relevant capacities into these two categories can be found in Fischer & Ravizza (1998), Brink (2013), Vargas (2013), among others.

\textsuperscript{11} Brink makes a distinction between the possession of normative competence and the performance of normative competence (Brink, 2013). Even though it might not be as straightforward as Brink puts it, the idea of this distinction, as well as the line of thinking to use it to reply to the challenge from situationism, can be found in Doris (2002), Vargas (2013), McKenna & Warmke (forthcoming), among others.
possession a more fundamental concept than exercise since the latter seems to be presupposed by the former.

The challenge from situationism towards moral responsibility is twofold. First, situationism casts doubt on the agent’s possession of normative competence. Second, it questions the regular exercise of normative competence. The former challenges moral responsibility by denying the categorical possibility of normative competence, that is, there is no such thing as normative competence. And if normative competence by definition precludes the influence of insubstantial situations on bringing about behavioral outcomes, the capacities to recognize and then act accordingly on moral reasons do not make sense when faced with the empirical evidence of situationism. The latter suggests that the agents fail to exercise normative competence on a regular basis. That is, even if human agency in general provides the basis to act, the agent fails to do so too often. If true, both claims put moral responsibility in jeopardy since the concept of moral responsibility is bonded with the possession and the exercise of normative competence. In the following two sections, I will discuss two approaches in reply, and try to show that neither serve as a sound reply unless we 1) clarify the relation between possession and exercise, and 2) lose certain concept of the morally responsible agent.

II. Argument from Possession

The possession of normative competence is key to the distinction between a responsible agent and a nonresponsible one. This is also supported by the common sense of culpability with the mentally disabled and moral cultivation with children. Thus, if situationism is true on the categorical impossibility of normative competence, then no one would count as responsible agent, and none is responsible for their behaviors. The response to the problem with possession comes handy since one can compromise the exercise of normative competence so as to keep the concept of possession thereof. Doris and Brink both have this idea of splitting exercise from possession:\footnote{1) Note that they have different definitions about normative competence, and they also adopt different theories of moral responsibility. These differences lead to different approaches arguing for and against situationism. I do not intend to discuss their differences for the present purpose, however, I try to focus on what I perceive as their similarities. 2) Brink suggests that Doris conflates possession with exercise rather than distinguishes them.}
Responsibility must be predicated on the possession, rather than the use, of such capacities...most situational findings describe patterns, albeit surprising patterns, in behavior. They do not demonstrate incapacity. (Brink, 2013: 8, 14)

The determination of whether a person manifested reasonable ethical awareness should not be determined by reference to the frequency with which failures of awareness occur. (Doris, 2002: 138)

Brink suggests that situationism at best shows that people's exercise of normative competence is not as expected, but the failure of exercise does not necessarily imply incapacity. Doris takes it even further to suggest that the frequency of failures is not forbidden either. This could easily slip to the claim that to be unaware of the morally salient information frequently does not entail the categorical impossibility of such capacity. Here I am not implying that either of them makes such claim as the exercise of normative competence of no importance, however, I admit that the approach to compromising the usage of normative competence on a regular basis is theoretically attractive, for, but not limited to, the following reason. Since the possession of normative competence is in a tight relation with one being a responsible agent, and it seems that the possession of certain capacity is more fundamental than the exercise thereof in the conceptual sense, in this way, even if situational factors can take credit for impugning on normative competence in the practical sense, they cannot dissolve one’s moral standing as a responsible agent.

But the problem with this approach is that it oversimplifies the relation between possession and exercise. For the purpose of clarification, here are three possible interpretations about the relation:

C1: The possession of normative competence does not entail the exercise of it every time it is appropriate to do so.

However, by the quotes I want to show the similarity in their tendencies to save something more general at the cost of degrading something more particular, despite their different aims of arguments.
C2: The exercise of normative competence this one time does not imply a genuine possession of it.

C3: The most reasonable relation between the possession of normative competence and the exercise of it is that the normatively competent agent regularly exercises normative competence.\(^\text{13}\)

C1 actually supports the view that even though the exercise of normative competence could admit situationism's challenge, this is the furthest the challenge could go, since the possession of normative competence allows for failures in exercise, there is no real challenge to the concept of possession, and in turn, to the concept of the morally responsible agent. However, unlike C1, C2 makes the distinction between possession and exercise a potential threat to the concept of possession, for it implies that exercise as a particular incident does not always refer to a more general concept or necessity of possession. Not only does C2 make the concept of possession less fundamental than it is assumed, if we take the stance on this interpretation of the relation between possession and exercise, it also adds force to situationism's challenge, for the empirical evidence might validify that there is a lack of exercise. C3 is a more flexible claim, it can be used to support each side since regularity as a standard for exercising normative competence is obscure nonetheless a practical issue on a daily basis, which adds to its obscurity when it comes to determining whether someone regularly exercises her normative competence or not.

Despite the conceptually ideal version of the relation being hard to testify, it is worth taking a look at the implication of such clarification on the concept of the morally responsible agent. If the concept is based on possession, then in addition to C2, situationism can challenge moral responsibility by showing that there is not the kind of responsible agent based on possession of normative competence, for situational factors, especially the insubstantial ones, can affect one's exercise to the extent that it might not

\(^{13}\) Regularity as a standard for responsibility-relevant capacities can be found in Fischer & Ravizza (1998), Nelkin (2005), and Brink (2013), among others.
be reasonable to perceive her as a possessor of normative competence, and in turn, a morally responsible agent anymore.\textsuperscript{14}

\section*{III. Argument from Exercise}
One way to respond to the problem from section 2 is to connect the exercise of normative competence and the concept of the responsible agent. For Doris and Brink, the concept of responsible agent is more correlated with possession of normative competence than with exercise. But if people do fail to exercise normative competence under the influence of insignificant situational factors, and the frequency of exercise is low enough to question the fundamental possession of normative competence among people, then the concept of responsible agent is unattractive, especially in the practical sense. In that regard, Vargas suggests another way thinking about situations, normative competence, and the responsible agent:

When (the sensitivities) are absent, or when they dip beneath a minimal threshold, the agent ceases to be a responsible agent, in that context. We need not suppose that if someone is a responsible agent at a given time and context, that he or she possesses that form of agency at all times across all contexts. In some contexts I will be a responsible agent, and in others not. Those might not be the same contexts in which you are a responsible agent. (Vargas, 2013: 20)

We can read two things from Vargas’ statement. First and foremost is that the concept of responsible agent does not stick to one person through time and/or situations. The concept of responsible agent is a role assigned to different people under the same circumstance, or to the same person under different circumstances. This is contrary to the concept of responsible agent presupposed by possession. Since possession is a unified concept, meaning that once someone meets certain required conditions, she is a possessor of something, and the thing is fixed and constantly held by her. In turn, it makes responsible agent a unified and fixed concept as well, meaning that

\textsuperscript{14}There is an ongoing discussion of whether insubstantial situational stimuli like implicit bias challenge moral responsibility, see discussions in Kelly & Roedder (2008), Holroyd (2012), Levy (2014), Brownstein (2016), Vargas (forthcoming), among others.
once assigned to a certain type of persons, there should be no exceptions. Meanwhile, situationism tries to show that the exceptions may be pervasive. Instead, if the concept of the responsible agent can be understood as a functional concept, alongside with exercise rather than possession, it makes the concept of morally responsible agent less threatened by situationism.

Second, the role of responsible agent is assigned based on whether the responsibility-relevant capacities are being exercised. As long as the exercise of normative competence is present, there is responsible agent. This is of merit for it helps reconnect the concept of morally responsible agent with exercise. That said, I am not implying that Vargas takes the stance to get rid of the concept of possession altogether, on the contrary, he acknowledges the conceptual necessity of possession for the kind of control required by moral responsibility (Vargas, 2013).

At this point, the concept of responsible agent is somewhat settled. At least one thing is clear: even if the empirical evidence may show a low frequency of exercise, it can not deprive us of the concept of responsible agent by depriving the categorical possibility of possession. However, the trajectory of this approach, even though not acknowledged by Vargas (2013), is easily committed to the abandoning of the concept of responsible agent based on possession. In other words, it degrades the concept of what “having something but rarely using it” means to the concept of responsible agent. It also brings about a difficulty unifying the concept of responsible agent correlated with exercise and that correlated with possession when faced with situationism. That is, these two kinds of concepts are exclusive to each other in the way that, in order to respond to situationism’s challenge on behalf of normative competence as the necessary condition for moral responsibility, either we admit that a concept of the responsible agent based on possession of normative competence is rather redundant in the practical sense, and we grant a concept of responsible agent linked with exercise. Or, we do not have a unifying concept of the morally responsible agent.

IV. Conclusion

To conclude, given that situationism challenges moral responsibility by denying the categorical possibility of possessing normative competence and by validifying the lack of exercise on a regular basis, neither the argument from possession nor the argument from exercise bear further examinations. For the argument from possession, the
approach to compromising exercise so as to deny incapacity and keep the concept of morally responsible agent is unsatisfactory, considering that short of exercise might in turn imply a possibility of incapacity. For the argument from exercise, though the concept of the responsible agent is understood as functional as a role assigned to whoever is exercising normative competence or whenever/wherever one is exercising it, this approach inevitably turns the concept of possession and a concept of responsible agent based on possession into redundancy.

References


Introduction

I have been thinking about the nature of free will for the past 25 years. In very general terms, my position has remained the same. But my views on the details have slowly evolved, in tandem with changes in the way I think about certain more basic metaphysical issues and about the nature of explanation. I will begin today by sketching out an account of freely willed action that represents my current way of thinking. I will then discuss three objections to my sort of view pressed recently by several authors. All of them center around what sorts of explanation are, or are not, possible for causally undetermined choices, and the implications for a person’s control over and moral responsibility for such choices.

I am a freedom-determinism incompatibilist: freedom of action is not compatible with its being true that, for every event, prior to the event there is an objective probability of 1 that the event will occur. If we have some measure of freedom, it must be true that for some of our choices, there was a non-zero prior probability that we would choose differently. While it is plausible that sometimes our choices—and particularly those we make ‘automatically,’ without deliberation and often without focal attention—are effectively causally determined, I also think that we make other choices where this is not the case, choices that are not causally determined, strictly or approximately.

For the sake of what follows, let us assume this conjunction of incompatibilism and causal indeterminism, the one an a priori philosophical thesis and the other an empirical matter. If it is indeed necessary for our sometimes acting freely and being morally responsible for the consequences of such actions that they have been causally undetermined, very plausibly it is not sufficient. What must certain of our actions be like, then, if we are to be free and responsible agents?

At this point, it is usual to distinguish event-causal and agent-causal approaches of freedom and to take a stand on which of these approaches is correct. Indeed, that problematic informs much of what I have written about free will, but I have come to think it is a mistake. The problem with it is not specific to freedom but stems from assumptions concerning causation quite generally. Event-causal analyses of causation, of some form or other, have been popular ever since David Hume, and especially
throughout the twentieth century. With a growing number of recent philosophers, I now think that the general identification of causes with events is an unfortunate legacy of Hume’s mistaken rejection of causal power and substance as basic metaphysical categories.

Deflationary neo-Humean metaphysics remains popular. But it is now being vigorously challenged by a broadly neo-Aristotelian alternative. On this alternative framework, the world is fundamentally a world of things/substances, not events. Events are the having of a power by an object. Accordingly, it is, in general, powerful particulars—objects—that exercise or manifest causal power. All causation is object or substance causation. To be sure, objects’ acting in the ways that they do have explanations: they reflect the causal powers that they have at the time, powers that are none other than (one or more) natural properties that they have, and (typically) reflect also the presence of external necessary manifestation conditions. (Only “typically”, since the phenomena of radioactive particle decay seems to involve no manifestation conditions.)

For example, two electrons, eddie and eleonore, mutually repel each other—that is, cause each other to accelerate along receding paths at a specific rate. They do so in virtue of their powers—that is, their common negative electric charge—and in the circumstance (necessary manifestation condition) of their being a certain distance apart with no significant countervailing forces.

I said that the “natural” properties of objects are causal powers. Here I refer to those fundamental, intrinsic features of objects that make for objective similarity. For an electron, such properties will be determinate magnitudes of charge, mass, spin, and magnetic momentum. For a human person, they will include such properties as believing, desiring, and intending various things.

In some possible cases, given the totality of properties had by an object and those of its environment, the effect is causally determined to occur. The conjunction of interacting powers yields a probability of 1 that the substance or substances will cause that very type of effect. In other possible cases, the effect is causally undetermined. Here, the exercise of more than one power is possible, and plausibly each of these distinct propensities is probable to some specific degree. Every indeterministic event is produced by a cause, though none is necessitated.

Let us then consider the case of a human person’s choice that is causally undetermined. Since we are trying to model free will, let us suppose a case where the
agent consciously decides between options while being aware of those factors, or at least many of them, that incline her towards each of the alternative possibilities she contemplates. These will include desires and aims she has, and beliefs concerning how she might attempt to realize those aims, her chances of success, and so on. In short, her relevant *motivating* reasons, or simply "reasons". As all behavior is motivated behavior, absent those or other motivating states, she would not have the power to choose the corresponding action. To be in a motivating state just is to have a power to act in at least some possible situation. Whether or not you are able to exercise that power at a particular time depends on the causal structure that results from the various powers in and around you, on what that total structure does or does not permit to occur. Where you are able in the circumstances to act in more than one way at some future time $t$, there will be evolving objective prior probabilities of your so choosing in each of these ways—evolving in response to your own deliberation, or to the lessened salience to you of an alternative, or to a relevant change in your environment that was itself not predetermined to occur.

What is it then to *act on* a particular reason, $R$? Put otherwise, what is it for the fact of your having had a motivational reason $R$ to explain your choosing and acting as you do? On the account I propose, your having reason $R$ explains your actual choice if it was among the factors that contributed positively to the prior probability of your making that choice throughout an interval that terminates in the choice itself. But this is consistent with its influence being wholly unconscious. So we might say that, you act *for* reason $R$ if you act *on* reason $R$ and you were consciously aware of having that reason at the time of the choice.

In summary: on the account I now favor, an agent acts freely and is a locus of moral responsibility (in the basic desert sense) if:

- he is a conscious and rational substance, with the general power to choose actions in accordance with his own motivations;
- he has at any given time more specific powers to choose particular courses of action that are grounded mostly in some of his motivational states, that run the gamut from momentary desires to long-range life aims;
- his choices are frequently causally undetermined just prior to the moment of choice;
- and he exercises that power in conscious awareness of some or all of the motivational
states that ground his power specifically to choose as he does.

I likewise contend that where these conditions obtain, the agent’s choosing as he does is rationally explained by the motivational states grounding his power so to choose. I spoke of that power’s being ‘mostly grounded in’ those states to simultaneously recognize that our conscious awareness of influences on us is not and need not be total yet also that too much significant, wholly sub-personal causal influence is freedom-diminishing or even undermining.

With this account on the table, I now consider three objections to it.

**Objection 1: Objective Probabilities and Control (Pereboom 61-2)**

In a recent book (2014), Derk Pereboom argues that supposing agent-causal power is governed by objective probabilities undermines responsibility-conferring control. He invites us to imagine that God creates beings who have two fundamental propensities, one for self-interest, and one for morality, and that these propensities are equal and inalterable. He notes that we should expect that over the course of a lifetime, roughly half of decisions involving a conflict of self-interest and morality would be self-interested, the other half moral. He takes it to be evident that in such a scenario, agents do not have the right sort of control over their choices. Anticipating (correctly) that some will object on the grounds that it’s the inalterability of the probabilities that is control-sapping, and this is an unnecessary feature, he then constructs a more complex case. In this revised case, God ‘adds a law’ that says that there is a 50% probability that at some point over each agent’s lifetime, the agent will in some way cause her own moral conversion, such that from that point onwards, all of her decisions will be moral. So we expect that about 50% of the agents will undergo such moral conversion, while the others do not. Here, too, he thinks, when we focus on the objective probabilities in play that make likely, in advance, a certain kind of complex distribution in choice outcomes, even though individual choices are left completely open, we doubt that the right sort of control is operative. (61-2)

Well, Pereboom simply describes his cases and pronounces his judgment. But I share to a significant degree his sense that there is something wrong in these scenarios, freedom-wise. And he himself puts his finger on the problem when discussing his first
scenario, but he fails to recognize (i) that it carries over to the second scenario, and to more complicated variations we could devise, and (iii) that it is wholly an artifact of his particular, non-mandatory way of modeling indeterminist choice. The problem is that Pereboom’s agents do not continuously change (in small or big ways) as a result of their individual choices. Pick an arbitrary point in my life. I have a choice between helping my daughter with her schoolwork or drinking beer with friends at the pub; I choose the pub. A couple days later, the particulars vary, but I have a structurally similar choice and again pursue the self-interested option. In real life, we’d think: careful! You’re starting to form a bad habit there. But not in Pereboom’s world: a string of ten self-interested (or moral) choices in a row is unlikely, but where it happens, I am not tilted in that direction, not even a bit. Nor do I learn from experience. Whatever reflection I do upon my weirdly balanced psychology makes no difference either. We might say more about what is freedom-undermining in this picture, but we’d first need to hear more about how the probabilities are psychologically embedded, such that they are impervious to alteration in any of these ways. Pereboom’s agents are very different from us in not being impacted by their choices, their experience of the consequences of choices, or occasional self-reflection. Supposing this is even possible, if we wanted to have a generally applicable account of responsibility-conferring control, we could augment our account by requiring such dynamical change in response to past choices, experiences, and reflection. There is no worry that human agents won’t meet such a condition.

Objection 2: Control, Luck, and the Absence of Contrastive Explanation (Nelkin 64-5; Levy 36, 43, 67)

On my view, an agent’s deciding to do A consists in the agent’s causing an executive intention to A. Where this is a free decision, there was a non-negligible prior probability that he would cause an intention to B instead. Dana Nelkin (2011) and Neil Levy (2011) both argue that such a scenario precludes anything more than a weak, defective kind of rational explanation of the choice the agent makes, and perhaps it allows for no explanation at all. Levy puts the charge as follows: if the agent decides to do A where B was really possible, then the pro-A set of motivations can explain only why A was being considered by the agent and why A was a causally open possibility. That ‘exhausts’ their explanatory power. To explain the choice of A itself, we would need to cite a factor that
contrastively explains why the agent chose A rather than B.

But where there is significant indeterminism, there cannot be contrastive explanation. A contrastive explanation would indicate why, given all the relevant influences at play, A was selected rather than B. But, ex hypothesi, those factors tug in different directions and no subset of factors dominate the others, as would be needed. It was simply a matter of ‘chance’, of good or bad ‘luck’, that A was chosen on this occasion, rather than its being something firmly under the agent’s control.

I think this objection is right in thinking that sheer inexplicability is inconsistent with agential control, but wrong in assuming that contrastive explanation is some kind of ‘gold standard’, or even the only standard. We do often desire and uncover different kinds of explanations. But when we ask for a contrastive explanation of the form, Why A rather than B?, we defeasibly presume that there is an explanatory relationship between fact (A) and ‘foil’ (not-B). That is, we presume that the occurrence of A and the non-occurrence of B can be given a unifying explanation. Even in a deterministic world, this will not be the case for every such pair of items, as not everything that happens is deeply explanatorily connected to everything else.

But non-contrastive explanations of particular phenomena are also possible, and in indeterministic scenarios of any kind, for certain purposes, they are the sole proper sort to seek. In general, an outcome of an indeterministic scenario is adequately explained by identifying the object or coordinated system of objects that produced the outcome, the features of the object that gave it the power to produce the outcome, and the total circumstance in virtue of which there was a nonzero probability of its doing so. This explanation is perfectly adequate even though there is no explanation of why this result was produced rather than that one, whose probability of occurring was likewise nonzero. Indeed, in such a scenario, we have an explanation of why there cannot be a correct contrastive explanation of the outcome, for every possible contrast— the very nature of an indeterministic causal agent precludes such explanations.

It perhaps helps to de-mystify indeterministic causal activity to think of deterministic causes as simply the limit cases of analogous indeterministic causes arranged on a continuum ordered by the strength of their antecedent probability to cause the actual outcome. Indeterministic causes (whether in basic physics or in human behavior) are not sub-standard causes, and there is nothing partial or otherwise defective about the noncontrastive explanations of them that accurate theoretical descriptions would
provide. Nothing pertaining to the target phenomena will be left out of good non-contrastive explanations of them: which events occur, how they are produced, and whether and why specific types of alternative events were possible, given the prior circumstances.

Now, Nelkin (and others) will say that, even if what I’ve just said is accepted, it’s hard to see how agents’ acting indeterministically “provides a unique kind of control” not exercised by otherwise similar agents acting deterministically. And if it does not, why think that indeterminism is a necessary condition on the kind of rational control that constitutes free will? While talk of “enhanced control” and its denial makes sense when contrasting “agent” and “event” causal accounts of freedom (a contrast the neo-Aristotelian metaphysics obliterates), it seems cloudy to me in the present framework. What we who believe that indeterminism is a necessary condition on freedom can clearly say is that indeterministic agents (but not their deterministic counterparts) are able to choose otherwise, taking into account all causal constraints under which they operate. They have, and their deterministic counterparts lack, the power to form different choices than the ones they do, for good or bad—not just a general power to do so, but the power to do so in the very circumstances they are in. And having this power is important to moral responsibility, whether or not we judge it appropriate to say it involves an ‘enhanced’ variety of control.

Objection 3: Impact of Non-chancy Luck on Choices and Character (Levy)

We enter the world with powerful and deep behavioral and attitudinal dispositions. Long before we mature to the point of making sophisticated, reflective choices, we are placed in environments that mold and add to those dispositions. As we ‘come of age’, these factors largely determine that Andy will choose from only a very limited range of options in any given situation, a range that will differ quite a bit from what Betty would consider in similar circumstances, and in many cases, these pre-given factors heavily influence our choices themselves, even if they do not causally determine all of them. These early choices and continuing contingencies of circumstance, in turn, will sharply circumscribe the options Andy considers at a more reflective stage, when we begin to hold him accountable for his actions.

A certain kind of free will/responsibility skeptic concludes from reflections such as these that debate over indeterminist agency is a sideshow: even if the sort of account I
have sketched describes some human behavior, it is not (nor could it ever be) enough. Since factors unchosen by Andy largely account for the kinds of deliberation and the overall pattern of outcomes of his mature choices, while he may have a small measure of freedom of choice, he lacks autonomy, and thereby deep moral responsibility.

In response, one must concede that responsibility for ‘shaping who I am’ and for the choices that ensue from this comes in degrees and, indeed, can only sensibly be measured within a limited scope of possibilities. We cannot hold Andy responsible for failing to consider an option entirely outside the range of his experience. And his responsibility for ignoring options which are within the range of his experience but which he has had precious little opportunity to consider as attractive is attenuated. In concrete cases, given limited information, we hazard rough guesses on these matters. When we are confronted with an individual who quite deliberately and unhesitatingly makes a grossly immoral choice—indeed, who seems not to even consider the obvious moral alternative—the question one needs to ask is this: was there a point earlier in her life when paths were open to her (ones for which at each step of the way she had some significant motivation to pursue, and which she recognized as having moral significance) such that had she taken them she would now be such as to see the force of the moral considerations at hand? How ‘difficult’ would it have been for her to pursue such a path? Our guesses about such matters are exceedingly rough, and rely on the assumption that a certain measure of rough moral sensitivity is had by most mature individuals. Absent compelling information to the contrary, then, we deem it appropriate to hold individuals responsible for their own moral indifference.

Moreover, we may consistently suppose that there are individuals whose basic choice-making capacities are just like ours, but who are largely insensitive to considerations of basic human decency through no fault of their own. And we can also accept that total responsibility for one’s choices and character is not just contingently lacking in human beings, it is an impossible notion: that would seem to require perfect indifference at the outset, or at least an openness to all possible courses of action. The coherence of that idea is very doubtful.

Even when we set aside extreme cases, involving horrendous childhood abuse, biologically-rooted deficiencies of moral cognition, and the like, we need to recognize that, as creatures of habit and settled psychological dispositions, our autonomy is a highly fragile thing. For some, early patterns of choices lead in foreseeable ways to
character and life outcomes in such a way that the agent reasonably thinks, "I have had a significant role to play in my becoming who I am." But for others, early choices cause them to fixate on certain goals, rashly, without their having foreseen or intended this consequence, and they come in later years to think of themselves as 'trapped' in a persona and pattern of life that feels imprisoning. diminishing rather than enhancing their future freedom over time. There is, in short, an element of luck in whether we fall into one or another of these categories; and the upshot seems to be that some come to have a greater measure of what we might term “diachronic moral autonomy” than others have, and this difference will not consistently track the degree of proper praiseworthiness or blameworthiness. This is a disquieting, even tragic reality. But I do not see that it warrants a generalized skepticism concerning moral responsibility.