

國立臺灣大學文學院哲學系

碩士論文

Department of Philosophy

College of Liberal Arts

National Taiwan University

Master's Thesis



一個新的自由意志與道德責任的來源相容論式觀點

A New Source Compatibilist View of Free Will and
Moral Responsibility

詹遠至

Chan, Yuan-Chih

指導教授：鄧敦民 博士

Advisor: Duen-Min Deng, Ph.D.

中華民國 113 年 6 月

June, 2024

國立臺灣大學碩士學位論文
口試委員會審定書

MASTER'S THESIS ACCEPTANCE CERTIFICATE
NATIONAL TAIWAN UNIVERSITY

A New Source Compatibilist View of Free Will and Moral
Responsibility

本論文係 詹遠至 (姓名) R10124005 (學號) 在國立臺灣大學 哲學系 (系/所/
學位學程) 完成之碩士學位論文，於民國 113年6月24日 承下列考試委員審查
通過及口試及格，特此證明。

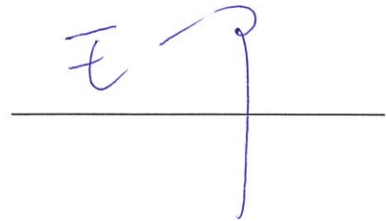
The undersigned, appointed by the Department / Graduate Institute of Philosophy
on 24 (date) June (month) 2024 (year) have examined a Master's Thesis entitled above
presented by Chan, Yuan-Chih (name) R10124005 (student ID) candidate and hereby certify that
it is worthy of acceptance.

口試委員 Oral examination committee:



(指導教授 Advisor)





系 (所、學位學程) 主管 Director:



誌謝



本篇論文得以完成，我首先要感謝我的妻子，朱倍玉小姐。她在我一開始煩惱是否要攻讀哲學碩士時鼓勵我；在我因撰寫論文而忙碌得焦頭爛額時，她包辦所有的家事；在我一度想要放棄時，她溫柔地支持我繼續堅持下去；在我每完成一個章節時，她共享我的喜悅並與我一同慶祝。若沒有她，本篇論文可能不會出現，或再花費一年才完成，或在中途被放棄，或是被孤單地完成。我也很感謝她總是認真聽我分享我的哲學研究，畢竟一般人可能不會對其感興趣。最後，我要感謝她總是不會缺席我的重要場合；例如我在台哲會研討會的發表，以及我的論文口試；她總是坐在台下認真聆聽，以及為我鼓掌。

接著，我要感謝我的指導教授，鄧敦民老師。鄧老師總是精準地為我推薦書籍；他推薦我閱讀的每一本書幾乎都在我撰寫論文的過程中提供非常重要的幫助。他也總是耐心地和我討論，並且配合我的節奏；當我要求每週都要 meeting 時，他也欣然答應。在討論時他總會提出許多精闢的建議，讓我想到原先沒考慮到的地方。

還有，我要感謝我的父母。他們對我在英文方面的栽培，讓我有能力以英文完成本篇論文，也讓我在閱讀英文文獻時非常輕鬆。當我最初選擇哲學系時，他們也完全支持，並沒有如世俗一般看輕哲學系。他們甚至還會以我的研究內容為榮；當我告訴他們我在研究自由意志與道德責任時，他們也認同我的研究的重要性。

最後，我要感謝周先捷老師與黃榮鑑老師。在碩士班的三年，兩位老師提供了無數的工作機會給我；我在他們提供的工作中領到合計超過百萬的薪資。

正是因為有兩位慷慨且樂於幫助學生的老師，我才能在不須擔憂經濟來源的狀況下完成學業。



Content



Chapter 1 Introduction: Free Will, Moral Responsibility, and Alternative Possibilities	1
1.1 Free Will and Alternative Possibilities	1
1.2 Free Will and Moral Responsibility	3
1.3 Moral Responsibility and Alternative Possibilities	4
Chapter 2 The Defenses of PAP, and the Replies to Them	8
2.1 The Flicker Defense	8
2.2 The Dilemma Defense	12
2.3 Blockage to Rescue?	14
2.3.1 The Blockage Cases	14
2.3.2 Objections to the Blockage Cases	17
2.3.3 The Last Hope for Blockage?	21
Chapter 3 A New Type of Blockage Case	28
3.1 The Illuminated Way	28
3.2 Developing a New Type of Blockage Case	29
3.3 PAP or PAP+?	33
3.4 Conclusion	39
Chapter 4 Toward the Source Compatibilism	41
4.1 Preliminaries	41

4.2 Pereboom’s Four-case Argument	43
4.3 Mele’s Zygote Argument	52
4.4 Replying to the Zygote Argument	55
4.4.1 Overlapping Control	56
4.4.2 The Adaptability of Control	61
4.4.3 Explaining Away the Intuition	67
4.4.4 Summing Up	70
4.5 Conclusion	71
Chapter 5 A New Source Compatibilism	72
5.1 Foreword	72
5.2 Ernie and the Reasons-responsive Theory	72
5.3 Lessons from Tim	74
5.4 A New Source Compatibilism	76
5.5 Closing Remarks	81
References	82



摘要



本篇論文的目的是探索一種新的來源相容論 (source compatibilism)。此計畫分為三個階段：建立來源式觀點、為相容論辯護，以及形式化我的理論。在第一部分，我發展了一個可以避免傳統「封鎖案例 (blockage cases)」以及其他「富蘭克福式案例 (Frankfurt-style cases)」所受到的挑戰的新版封鎖案例。我利用這個新版封鎖案例來否定一個版本的 PAP：PAP+；並以此論證來源式觀點的合理性。在第二部分，做為對相容論的辯護，我發展了一個對 Alfred R. Mele 的受精卵論證 (zygote argument) 的回應。我論證就算一個行為者是被一個崇高存在有意圖地創造的，他仍然可以自由且須負道德責任地行動。在最後部分，立基於我在前兩部分的發現，我提出了一種新的來源相容論；它具有一個核心條件：實際控制條件 (actual-control condition)。

Abstract



This thesis explores a new type of source compatibilism. The project is divided into three main parts: establishing the source view, defending compatibilism, and formulating my theory. In the first part, I develop a new type of blockage case which does not suffer from objections that traditional blockage cases and Frankfurt-style cases in general face. By this new blockage case, I argue for the falsity of a version of PAP, PAP+, and thus establish the plausibility of the source view. In the second part, I develop a reply to Alfred R. Mele’s zygote argument, as a defense of compatibilism. I argue that even if an agent were intentionally designed by a supreme being, the agent could still be free and morally responsible for his actions. In the final part, I propose a new type of source compatibilism with a core condition, the *actual-control condition*, which is based on my findings from the first two parts.

Chapter 1



Introduction: Free Will, Moral Responsibility, and Alternative Possibilities

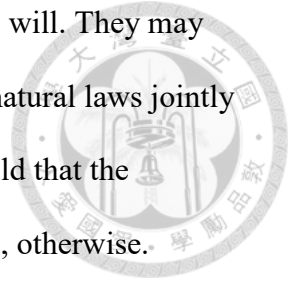
1.1 Free Will and Alternative Possibilities

For a person to have free will, it must be the case that he has certain *control* over his intentions, decisions, and actions. That is to say, those mentioned above must be “up to” this particular person. Intuitively, we link this important kind of control to the ability to access *alternative possibilities*. Being able to access alternative possibilities means that one is able to do otherwise than he actually did. For example, this morning I chose to ride a bike to school. However, it was totally open to me that I could have chosen to take a bus to school instead. There were some alternative possibilities accessible to me when I chose to ride my bike, one of which was taking the bus.

To many philosophers, this accessibility to alternative possibilities is crucial to free will. They believe that for one to be a free agent, he must have the ability to act otherwise than he actually did. In this thesis, I call this group of philosophers “leeway theorists.”¹ Whether being a compatibilist or an incompatibilist, the leeway theorists

¹ I borrow this term from Derk Pereboom (2001). The leeway theorists include compatibilists and also incompatibilists. Prominent examples for leeway compatibilists are classical compatibilists like A. J. Ayer (1954), and those for leeway incompatibilists are Peter van Inwagen (1983) and Robert Kane (1996).

believe that the ability to do otherwise at least partially grounds free will. They may differ in whether determinism, the thesis that past world states and natural laws jointly determine future world states, threatens our free will, but they all hold that the important *control* pertinent to free will is the ability to act, or to will, otherwise.



Hence, if determinism doesn't rob us of our free will, that is because we can keep our ability to act otherwise under determinism, and if determinism does take away our free will, it does so by depriving us of our ability to act otherwise.

Intuitive as it seems, I believe that the leeway theorists are at least in some sense mistaken. There is another group of philosophers that I side with; I shall call them "source theorists."² The source theorists believe that the crucial *control* pertinent to free will is not the ability to act otherwise, but to be the *ultimate source* of your action. For example, when I chose to ride my bike to school this morning, I was free not because I could have chosen to take a bus, but because the source of this choice was me. The choice was not the result of some coercion or manipulation. It was caused by my deliberation. As we can see, the source theorists focus not on the possible ways things could have been, but on how things *actually* were. They look into the actual causal sequence of an action to determine whether this action is freely done.

In this thesis, I will try to defend the view that the crucial *control* pertinent to free will is not the ability to act otherwise, but to be the ultimate source of one's action, and that free will is compatible with determinism. In short, I would like to provide an argument for source compatibilism.

² This term was from Michael Mckenna (2001). Source compatibilism is also famously called "actual sequence view." Prominent proponents of it are John Martin Fischer and Mark Ravizza (1998).

1.2 Free Will and Moral Responsibility



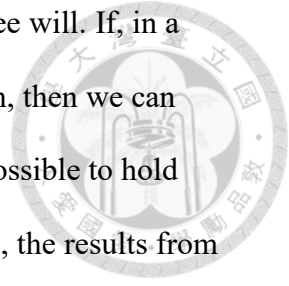
Before starting to discuss the reasons for preferring the source view over the leeway view, we need one more bridging step: connecting free will with moral responsibility.

The connection between free will and moral responsibility is straightforward. We commonly think that it would be unfair or improper to hold someone morally responsible for an action, if he did not have the relevant control over his behavior, or that action was not up to him. We tend to take back our praise or blame toward someone, if we find out that he could not control himself over his decisions or actions. I believe that this folk theory of moral responsibility is accurate. It captures a *necessary*, though not sufficient, condition for moral responsibility. Hence, I will follow Alfred R. Mele (2006, 17) in taking free will as “*the strongest control condition required for moral responsibility.*”³ That is to say, having free will is a necessary condition for being morally responsible for one’s action. It is not possible that an agent is morally responsible for his action but lacks free will. I think this definition of free will is plausible, since I believe the sense of *control* in free will is the same sense of *control* that is required in moral responsibility: we only hold people morally responsible for actions which are *up to* them, *up to* them in the sense that they use their free will in performing those actions.

Making clear that free will is a necessary condition for moral responsibility is very useful in the debate between the leeway theorists and the source theorists. The

³ Susan Wolf (1990, 3-4) and Fischer (1994, 3) also shared the view that free will is the necessary condition for moral responsibility. Harry Frankfurt (1971, 18-20) and Eleonore Stump (1988) disagree, but only in the cases of derivative responsibility. Here I am restricting my discussion to direct responsibility.

judgments on moral responsibility can serve as the litmus test for free will. If, in a scenario, we judge that an agent is morally responsible for his action, then we can infer from this judgment that this agent has free will, since it is impossible to hold agents that lack free will morally responsible for their action. Hence, the results from the analyses of the theories of free will had better coincide with our judgments of moral responsibility. If according to one theory of free will, an agent is *not* free when doing something, however, based on some convincing arguments, this agent *is* morally responsible for his action, then this theory of free will is in big trouble. If we want free will to be the control that is necessary in cases of moral responsibility, then our theory of free will has to give the same verdict as that of moral responsibility.⁴



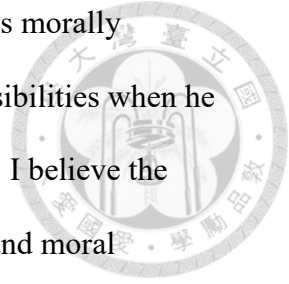
1.3 Moral Responsibility and Alternative Possibilities

Based on the connection between free will and moral responsibility, my strategy for supporting the source theory is to show that the leeway theory of free will does not give the right verdict in cases of moral responsibility. Hence, it does not capture the *control* that is necessary for moral responsibility, namely, free will. Furthermore, I will try to show that the source theory of free will does not face the same problem as the leeway theory. Thus, the source theory better captures the *control* that is crucial in moral responsibility; that is to say, the source theory is a better theory of free will.

In a ground-breaking article, Harry Frankfurt (1969) developed a kind of thought experiment now famously called the “Frankfurt-style cases (Frankases for short).”

⁴ To be clear, in cases of *lack of* moral responsibility, the judgments of free will and moral responsibility can be different. Free will is not a sufficient condition for moral responsibility, hence, there can be cases in which the agent is not morally responsible for his action even though he freely does it.

The Frankases are intended to show that it is possible that an agent is morally responsible for his actions even though he lacks any alternative possibilities when he acts. In this thesis, I will appeal to the Frankases to defend my view. I believe the Frankases illuminate the right path toward the analysis of free will and moral responsibility; that is, the ability to do otherwise is *not* the crucial control pertinent to them.



First, let's take a look at what a Frankase looks like:

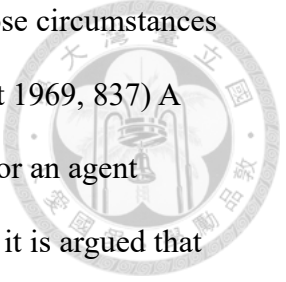
Suppose someone—Black, let us say—wants Jones₄ to perform a certain action. Black is prepared to go to considerable lengths to get his way, but he prefers to avoid showing his hand unnecessarily. So he waits until Jones₄ is about to make up his mind what to do, and he does nothing unless it is clear to him (Black is an excellent judge of such things) that Jones₄ is going to decide to do something other than what he wants him to do. If it does become clear that Jones₄ is going to decide to do something else, Black takes effective steps to ensure that Jones₄ decides to do, and that he does do, what he wants him to do. Whatever Jones₄'s initial preferences and inclinations, then, Black will have his way.⁵ (Frankfurt 1969, 835)

At the end, Jones really decides on his own to do, and does do, what Black wants him to do. Hence, Black does not have to show his hands, and does nothing except monitoring Jones's decision process.

Proponents of Frankases try to build “IRR-circumstances”⁶ which “make it

⁵ For a more elaborate Frankase, see Fischer (2006, 38).

⁶ This term was from David Widerker (2000).



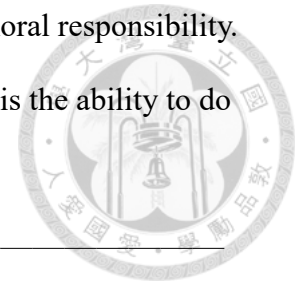
impossible for a person to avoid performing some action without those circumstances in any way bringing it about that he performs that action.” (Frankfurt 1969, 837) A successful IRR-circumstance eliminates all alternative possibilities for an agent without causing the action of that agent.⁷ In the IRR-circumstances, it is argued that the agent is nevertheless morally responsible for his actions. It is because the fact that the agent has no alternative possibilities “played no role at all in leading him to act as he did. He would have acted the same even if it had not been a fact. Indeed, everything happened just as it would have happened without Black's presence in the situation and without his readiness to intrude into it.” (836) That is to say, Jones chose to do what he did totally by his own reasons, not being affected by Black at all. All that Black had done was to monitor and to be prepared to intervene if necessary. Black was a “counterfactual intervener”—he would only intervene in counterfactual scenarios, but in the actual scenario he did not. Hence, Jones should take full responsibility for his actions. After all, even if Black was absent, Jones would have done exactly the same thing.

If IRR-circumstances are possible, and agents in IRR-circumstances are morally responsible for their actions, then an old and seemingly unimpeachable principle is mistaken. That is the “principle of alternative possibilities (PAP).” PAP says that a person is morally responsible for what he has done only if he could have done otherwise. This principle places the ability to do otherwise as a necessary condition for moral responsibility. Hence, it is crucial to the leeway theory of free will. If we define free will as the necessary control condition for moral responsibility, and combine this definition with PAP, then we get the leeway theory:

⁷ As we will see later, my position is that a successful IRR-circumstance does not have to, and should not, eliminate *all* alternative possibilities for an agent.

Definition_{fw}: Free will is the necessary control condition for moral responsibility.

PAP: The necessary control condition for moral responsibility is the ability to do otherwise.



Hence, LEE: Free will is the ability to do otherwise.

Such core principle for the leeway theory is undermined by the Frankases because, the agents in the Frankases do not have the ability to do otherwise, they can only do what they actually did, nonetheless they are morally responsible for their actions. That is to say, it is possible that there are situations in which agents are morally responsible for what they have done but lack any alternative possibilities. If the Frankases are successful, PAP is false.

In contrast to the leeway theory, the source theory has no problem in explaining the moral responsibility of Jones. Jones's decisions and actions are the result of Jones's deliberation. Jones is the source of them. Whether or not Black is lurking behind Jones, as long as Jones makes decisions on his own reasons, not being coerced or manipulated to do so, then he should take full responsibility for his decision and actions. Hence, "being the ultimate source of one's actions" has no problem in serving as the necessary control condition for moral responsibility. The source theory gives the right verdict in the Frankases.

Chapter 2



The Defenses of PAP, and the Replies to Them

2.1 The Flicker Defense

Although the Frankases are powerful counterexamples to PAP, they themselves face some objections. The Frankases have two crucial features, and both of them seem to be necessary to the Frankases: first, the agent in the case does not have any alternative possibilities when he acts; second, he is morally responsible for his actions. The objections refuse that the Frankases have successfully established these two features, hence claim that they are not counterexamples to PAP. The first objection, called “the flicker defense,”⁸ focuses on the former feature.

The flicker defense against the Frankases claims that there are still some alternative possibilities left in the scenario of the Frankases, hence PAP is not false. Recall the original Frankase designed by Frankfurt. In this scenario, although Jones₄ cannot actually make any decisions, or do anything other than what Black wants due to Black’s readiness to intervene, it is at least possible for him to show an inclination that he is going to make decisions that are against Black’s will. This showing of an inclination that he does not actually show, called “prior sign” by many, seems to be a necessary element in any Frankases since Black needs such sign to decide that he is

⁸ The flicker defense was originally developed by van Inwagen (1983, 161-82). Its name was later made famous by Fischer (1994).

going to show his hands. Hence, it seems that Jones₄ has a “flicker of freedom” here: he can show this inclination or not show it. At the point before making the decision, Jones₄ does have alternative possibilities. Here, the proponents of PAP claim that the presence of these alternative possibilities explains why Jones₄ is free, and why he is morally responsible for the decision that he actually makes and his consequent actions.

However, as had been pointed out in Fischer (1994), this kind of flicker defense can be easily solved by rewriting the case. We will only have to rewrite the case so that Jones₄ cannot even voluntarily show the inclination that he is going to make decisions that are against Black’s will. For example, let’s say that this prior sign that Black needs is Jones₄’s involuntary blush. If he is going to make any decisions that are against Black’s wish, he will first involuntarily blush. Jones₄’s blush is the only prior sign that Black needs since once Jones₄ blushes, Black can know in advance that Jones₄ is not going to follow his wish, and he will know that he has to step in and make sure that Jones₄ make the right decision and do what he wants. If we rewrite the Frankase in this way, then Jones₄ cannot even voluntarily show any inclinations other than the one he actually shows. In this way, we seem to extinguish all the flickers of freedom.

Nonetheless, the proponents of PAP have another move. In McKenna (1997), he argued that although there are no “normal” alternative possibilities presence in the revised case, there are still some alternative possibilities left for Jones₄: to blush or not to blush. McKenna claimed that this teeny-tiny alternative possibility makes all the difference. Although Jones₄ cannot voluntarily choose to blush or not, it is still the case that if Jones₄ blushed, Black will intervene and force Jones₄ to decide to do what he wants Jones₄ to do; if Jones₄ did not blush, then Jones₄ will decide on his own to

do the same action. That is to say, for Jones₄, blushing or not means to “escape the authorship” of his action or not, and this is very significant. Given that Jones₄ indeed has the alternative possibility of blushing, we can say that Jones₄ could avoid being the author of his action, hence avoid being morally responsible for it. However, in what transpires, he does not escape the authorship of his action that he could escape, and that explains his moral responsibility.⁹

In reply, Fischer developed the notion of “robust alternatives.” In his (1999), he argued that the control that the proponents of PAP are looking for cannot be provided by the mere existence of any sorts of alternative possibilities. Only the accessibility to robust alternatives can provide the desired control for the proponents of PAP. Robust alternatives are the alternatives that we believe we have in our everyday lives before learning philosophy. They are alternatives that we can voluntarily choose. We weight them consciously or unconsciously, and make a choice between them. Moreover, they are morally significant alternatives.

For example, when I am pointed by a gun and forced to rob a bank, rather than robbing the bank, I could also shut myself down and lose consciousness so that I will not have to rob the bank. This alternative is morally significant since I could absolve myself from blame by doing so. However, I cannot voluntarily choose to shut myself down, hence it is not a robust alternative. Consider another pair of alternatives, instead of taking five steps to the bank counter, I could take six steps to there. I can voluntarily choose between them, but it is not morally significant since how many steps I take does not affect my moral responsibility. Hence, they are still not robust alternatives.

⁹ Here I am using interchangeably PAP and some related principles focusing on avoidability developed by Keith Wyma (1997) and Michael Otsuka (1998).

Given what has been said, in Jones₄'s case, blushing is not a robust alternative, since that is not what Jones₄ could voluntarily choose to do. That is to say, although he could avoid authorship of his actions and blame by blushing, he could only do so accidentally. It would be only by luck if he avoids moral responsibility by blushing.

Fischer argued, if the proponents of PAP want to ground his moral responsibility on the existence of alternative possibilities, there have to be alternatives like voluntarily making another decision, or voluntarily acting otherwise. However, none of these are present in the revised case. The possibility of involuntarily blushing is not enough to ground his moral responsibility. What is important is not that whether Jones₄ could avoid moral responsibility by blushing, but whether he could choose to do so voluntarily. As a result, the flicker defense fails.

Nevertheless, I believe that the flicker defense cannot be dismissed so easily. To use Fischer's idea to respond to the flicker defense, we need a Frankase in which no robust alternatives are present. That is to say, we can only accept non-robust alternatives in our case, and these non-robust alternatives are provided by some involuntary prior signs of the agent, like the blushing in Jones₄'s case. However, we can build a successful robust-alternatives-free Frankase, only if we assume a deterministic relation between the absence of the involuntary prior sign and the desired decision or action. Let me elaborate. To eliminate all robust alternatives, we need Black, the counterfactual intervener, to be able to intervene before any voluntary signs can be shown. That is because once a voluntary sign is shown, like the forming of an intention that the agent is going to make such and such decisions, a robust alternative would be available for the agent. The alternative would be robust since the agent could voluntarily choose to show that sign or not, and this is what we cannot allow. Hence, we need an involuntary prior sign that its absence can guarantee no

undesired voluntary choice will be made. As long as the prior sign is not shown, the counterfactual intervener can be sure that it is impossible for the agent to make any voluntary choices that are against his wish. Let's say that Black wants Jones₄ to make decision D at the time t_2 , and if Jones₄ is going to make any decisions other than D at t_2 , he would involuntarily blush at t_1 . Hence, if at t_1 Jones₄ does not blush, then Black can be sure that Jones₄ will make D at t_2 . However, here comes the problem: for this case to work, we have to assume that the absence of blushing at t_1 is causally sufficient for Jones₄'s making D at t_2 . This leads us to the second objection that the Frankases face.

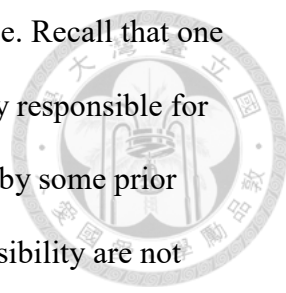
2.2 The Dilemma Defense

Given what has been said above, I believe that responding to the flicker defense by adding an involuntary prior sign to the case gives rise to another problem, the dilemma defense.¹⁰ The dilemma defense is another way to defend PAP. It claims that the Frankases face a dilemma.

The first horn of the dilemma is what has been discussed in the last paragraph of 2.1: we need an involuntary prior sign that its absence at t_1 is causally sufficient for the agent's desired actions at t_2 . However, if the absence of the prior sign at t_1 is causally sufficient for the agent's actions at t_2 , it seems that we can say the absence of the prior sign at t_1 *causally determines* the agent's actions at t_2 , in the sense that given the prior sign's absence at t_1 , the desired actions are determined to be performed at t_2 .

If the agent's actions are causally determined by the absence of the involuntary

¹⁰ The dilemma defense was developed by Widerker (1995), Carl Ginet (1996), and Kane (1996).



prior sign, then we, the proponents of the Frankase, are in big trouble. Recall that one of the necessary features of the Frankases is that the agent is morally responsible for his actions. However, if the agent's actions are causally determined by some prior sign out of his control, then the incompatibilists about moral responsibility are not going to agree that he is morally responsible for his actions. Our claim that "the agent is morally responsible for his actions even though they are causally determined by some factors out of his control" would be begging the question against the incompatibilists about moral responsibility. Hence, it seems that in replying to the flicker defense, although Fischer has successfully eliminated all robust alternatives, he nevertheless loses the ground to persuade the incompatibilists about moral responsibility that the agent is indeed morally responsible for his actions.

Now let's turn to the other horn of the dilemma. What if the absence of the prior sign does not causally determine the agent's action? It seems that we are still in trouble. If given the absence of the prior sign at t_1 , it is still possible that the agent does not perform the desired actions at t_2 , then Black, the counterfactual intervener, would not be able to decide whether he should intervene or not. If Black does not intervene, like what actually transpires, then it is actually possible that the agent does not perform the actions that he wants. That is to say, in this case the agent would have robust alternative possibilities when he acts, since he could perform other voluntary actions. If Black intervenes anyway, then the agent would not be morally responsible for his actions. Either ways, we do not have a counterexample to PAP.

To sum up, the Frankases seem to have gone into a dead end: to reply to the flicker defense, we need to replace the voluntary prior sign with the involuntary one; however, either the involuntary prior sign causally determines the agent's actions, or it does not; no matter which way we choose, we cannot falsify PAP.



2.3 Blockage to Rescue?

Now, let's see how the proponents of the Frankases save their cases from the above objections. To avoid the dilemma defense, many theorists have tried to revise their cases. In the literature, there are four types of modified Frankases: the no-prior-sign cases,¹¹ the blockage cases,¹² the internal-sign cases,¹³ and the necessary-condition cases.^{14,15} These modified cases either try to avoid the illicit presupposition that the prior sign causally determines the agent's actions, or do not use prior signs at all. In what follows, I will focus my discussion on the blockage cases since they are the theme of this thesis.

2.3.1 The Blockage Cases

The blockage cases were firstly developed by Hunt.¹⁶ In his (2000), Hunt proposed three kinds of blockage cases; I label them as the actual blockage case, the backward time travel case, and the foreknowledge case.

The actual blockage case is a case in which all the neural pathways of the agent are blocked in advance except the one that is actually activated. In this case, Hunt assumes that the agent is deliberating in a normal fashion, without any counterfactual interveners lurking behind him. Everything is normal except that a mechanism is

¹¹ See Mele and David Robb (1998).

¹² See David Hunt (2000).

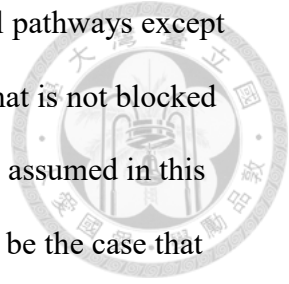
¹³ See Stump (1996) and Ishtiyaque Haji (1998, 38-9).

¹⁴ See Pereboom (2000) and (2001, 18-28), also Hunt (2005).

¹⁵ I borrow these names from McKenna and Widerker (2003, 9-10).

¹⁶ Actually Mele and Robb's no-prior-sign case (1998) involves some kind of blockage mechanism, but Hunt made blockage as the core element of his cases.

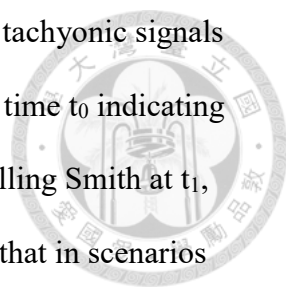
implanted in the agent's brain. This mechanism blocks all the neural pathways except one in advance. Due to sheer coincidence, the one neural pathway that is not blocked is the one taken by the agent. In addition, no causal determination is assumed in this case. The agent is deliberating indeterministically. It just happens to be the case that the neural pathways not activated are all blocked in advance. (217-8)



Now let's turn to the backward time travel case. This kind of blockage case is especially interesting. Let's see how Hunt described it in the original text:

Suppose, however, that the counterfactual mechanism is omitted from the picture, leaving only the indicators for Jones's actual action. So posit an actual sequence satisfying PAP¹⁷ (including causal indeterminism and/or agent-causation at all those junctures where libertarians would require it for Jones's moral responsibility), with the further stipulation that the sequence is continuously emitting tachyonic signals (consisting of particles which travel backward in time) encoding complete and unambiguous information about Jones's current states. Then Jones's killing of Smith, along with all the preliminaries to it (such as deciding to kill Smith) that are relevant to his moral responsibility, will be preceded by tachyonic indicators given which Jones's actions are unavoidable. But these indicators do not causally determine Jones's future action(s); they are not even part of the actual sequence leading up to the murder (they come "later" in the sequence - not temporally, but explanatorily). (218-9)

¹⁷ PAP includes all the necessary conditions for moral responsibility other than PAP, for example, the epistemic condition.



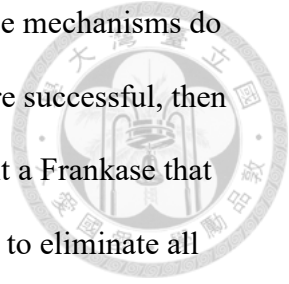
In this case, Jones's killing of Smith is rendered unavoidable by the tachyonic signals that can travel back in time. The showing of the signals at an earlier time t_0 indicating that Jones will kill Smith at a later time t_1 is sufficient for Jones's killing Smith at t_1 , given the nature of the tachyonic signals. In addition, Hunt claimed that in scenarios in which the conditions rendering an action unavoidable are not the conditions actually contributing to that action, with PAP satisfied,¹⁸ we would have the intuition that the agent is morally responsible for his action. Since the tachyonic signals do not contribute to Jones's killing Smith, but the other way around, Hunt claimed that we have the intuition that Jones is responsible for killing Smith. In conclusion, we seem to have a Frankase in which the action is unavoidable, but the agent is still morally responsible for that action. I believe that the backward time travel case deserves more discussion than it has received. In very soon, we will come back and talk more about this kind of blockage case.

Finally, the foreknowledge case involves an inerrant predictor, let's say Black. Black can predict the agent's decisions and actions correctly, hence renders them unavoidable. Just as the above cases, no causal determination is needed. We only need an agent that has libertarian free will, and an inerrant predictor that has infallible foreknowledge. (219-20) As we can see, the foreknowledge case is similar to the backward time travel case in structure: the actions in both kind of cases are rendered unavoidable by some kind of "sign" indicating future states that is present or knowable at an earlier time.

In these three blockage cases, some important features are shared: all alternative possibilities are eliminated not by a counterfactual intervener, but by an actual mechanism; also, we still have the same intuition as in the original Frankase that the

¹⁸ This kind of scenario is actually the IRR-circumstance.

agent is morally responsible for his decisions and actions, since these mechanisms do not cause the agent to behave in these ways. If the blockage cases are successful, then we have found a way to sidestep the dilemma defense. We have built a Frankase that does not assume causal determination, and at the same time manage to eliminate all the alternative possibilities of the agent.



2.3.2 Objections to the Blockage Cases

Unfortunately, the blockage cases face strong objections. Theorists including Fischer, Pereboom, Kane, and Widerker have all discussed the difficulties of the blockage cases.¹⁹ The main difficulty of them is that in actually eliminating all the alternative possibilities of the agent, the blockage cases come too close to causal determination.

This is a good place to make clearer what the “deterministic horn” of the dilemma defense is actually claiming. We already know that if the agent in our case acts deterministically, then we are begging the question against the incompatibilists about moral responsibility. But, what exactly is meant by “acting deterministically?” Consider the original Frankase. Even if we set aside the prior sign, we can still say that Jones₄'s deliberation process and Black's readiness to intervene jointly *determine* Jones₄'s resulting actions, since Jones₄ could not do otherwise. However, in this case, we would not say that Jones₄ *acts deterministically*, since Black actually does nothing and that Jones₄ makes the decision all by his own, hence the fashion in which Jones₄ acts can still be *indeterministic*. In this case, no question is begged against the incompatibilists, since there is no causal determination in what actually happens. The

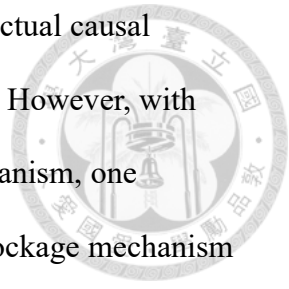
¹⁹ See Fischer (1999), Pereboom (2000), Kane (2000), (2003), and Widerker (2000), (2003).

crucial point here is that Black is not part of the *actual causal sequence*; Black contributes nothing to Jones⁴'s resulting actions. Hence, it is now clear that the causal determination that matters to the incompatibilists about moral responsibility, is the one within the scope of the *actual causal sequence*. That is to say, as long as the actual causal sequence before the resulting action does not determine one single action, we have the ground to persuade the incompatibilists that our agent is morally responsible for his action. The reason why the original Frankase falls into the dilemma defense is that its causal determination happens between the prior sign and the resulting actions, and that presumably the prior sign is part of the actual causal sequence. In conclusion, the lesson taught by the “deterministic horn” of the dilemma defense is that a successful Frankase should avoid causal determination between what actually causes the action, and the action itself.

Now, we know that whether the blockage cases have successfully escaped from the dilemma depends on whether the alleged causal determination is within the actual causal sequence or not. If the blockage mechanism is like Black, a mere counterfactual intervener, then there would be no causal determination within the actual causal sequence. However, if the blockage mechanism is part of the actual causal sequence, and that one single action is determined by it, then we would still be begging the question against the incompatibilists about moral responsibility.

In Hunt's original cases, he supposed that the fashion in which the agent acts could be indeterministic, and that the blockage mechanism does not feature in the actual causal sequence leading to the action. Hence, we have every reason to have the same intuition about moral responsibility as in the original Frankase. However, many theorists have forcefully pointed out that the blockage mechanism does engage in the actual causal sequence, thus turn the process from indeterministic to deterministic. To

see this, if the agent acts indeterministically, then it means that the actual causal sequence leading to the action does not determine one single action. However, with the participation in the actual causal sequence of the blockage mechanism, one specific action is determined. Pereboom vividly showed how the blockage mechanism can make a difference to the actual causal sequence by considering a two-situation case:



Situation C: A spherical atom is falling downward through space, with a certain velocity and acceleration. Its actual causal history is indeterministic because at any time the atom can be subject to an uncaused swerve. Suppose that the atom can swerve in any direction other than upwards. In actual fact, from t_1 to t_2 it does not swerve.

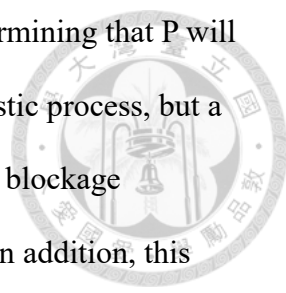
Situation D: The case is the identical to C, except that the atom is falling downward through a straight and vertically oriented tube whose interior surface is made of frictionless material, and whose interior is precisely wide enough to accommodate the atom. The atom would not have swerved during this time interval, and the trajectory, velocity, and acceleration of the atom from t_1 to t_2 are precisely what they are in C. (2000, 127)

The actual causal sequences in situation C and D seem to be the same. However, after close examination we can see that they are different, and this difference is contributed by the presence of the tube (the blockage mechanism). The tube makes the falling atom unable to go through any swerves that it originally could go through. Although the results are the same, the causal sequences are not: in situation C, the falling process singly causes the outcome; but in situation D, the falling process and the tube

jointly cause the outcome. The tube contributes to the causal sequence by helping decide the (only) direction that the atom could fall, thus making falling downward straightly the determined outcome. If this is not enough, later I will provide a more detailed argument about the claim that the blockage mechanism and the agent jointly cause the outcome, thus the blockage mechanism is part of the actual causal sequence.

The falling atom case is analogous to the actual blockage case. In the actual blockage case, the actual causal sequence of the blockage scenario is different from that of the normal scenario. The blockage mechanism helps decide the (only) neural pathway that is going to be activated, thus making the deliberation process deterministic. Those still not persuaded may wonder: but the blockage mechanism does nothing to the deliberation process! Even without the mechanism, the exact same neural pathway would still be activated. However, I think this way of construing the actual blockage case is mistaken. Here, consider my argument that the blockage mechanism is part of the actual causal sequence, and that this makes the actual causal sequence deterministic.

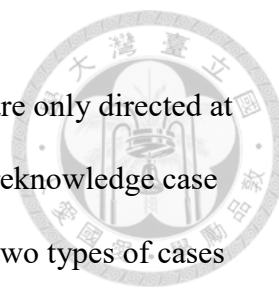
The following way of construing an actual blockage case could be misleading: first, imagine an indeterministic deliberation process P which ultimately leads to the decision D; and second, block every neural pathway that is not used in P. In this way of construing, we tend to think that P has *already* led to D, and we are just blocking all the roads not taken, hence P is not affected at all. However, this is wrong. What we actually have in the beginning is just an indeterministic process P that once starts running, could end up leading to decision D, E, F, or X, or what have you. We could not first let P finish in D, and then do the blocking, since in that case P would have plenty of alternative possibilities while running. We have to set the blockage *in advance*, not knowing which decision P will lead to. Hence, in blocking all the neural



pathways not used *if* P is going to lead to D, we are actually predetermining that P will lead to D, rather than E, F, or X. Now, P is no longer an indeterministic process, but a deterministic one. Since, the beginning state of P, equipped with the blockage mechanism, determines that D will be the outcome of this process. In addition, this result is not singly caused by P, since the blockage mechanism does contribute to it. We could not take away the blockage mechanism, like we take away Black in the original Frankase, and expect that everything would be the same. Since, the blockage mechanism is the crucial factor that help decide in the first place that D will be the outcome. We could subtract Black from the original Frankase without affecting the result, since Black has done nothing to Jones₄'s deliberation process. However, in the actual blockage case, the blockage mechanism has done *a lot* to P. Thus, we could not subtract the blockage mechanism from the scenario and expect that the result would be exactly the same. In conclusion, I think the blockage mechanism is part of the actual causal sequence that lead to D, since it is a partial cause of D. Thus, the causal determination in the actual blockage case is the one within the actual causal sequence, which begs the question against the incompatibilists about moral responsibility.

The objections discussed above show that the actual blockage case has not yet escaped from the “deterministic horn” of the dilemma defense. The actual causal sequence of the scenario is a deterministic process. Hence, claiming that the agent in this scenario is morally responsible for the actions led by this process, is begging the question against the incompatibilists about moral responsibility. These considerations are what probably make Hunt concede that “blockage is probably not the magic bullet for which PAP’s critics are looking” in his (2005).

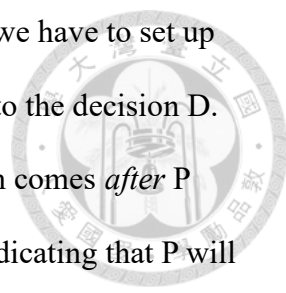
2.3.3 The Last Hope for Blockage?



However, readers may have noticed that the above objections are only directed at the actual blockage case. The backward time travel case, and the foreknowledge case are very different from the actual blockage case in structure. These two types of cases do not involve blockage mechanisms that are *added* to the actual causal sequences; rather, their blockage is done by factors *from* the actual sequences. In the backward time travel case, the blockage is done by the tachyonic signals emitted from the actual causal sequence; and in the foreknowledge case, the blockage is done by the knowledge of the inerrant predictor, which in turn comes from the future facts of the actual causal sequence. In both types of cases, the actual causal sequences are rendered unavoidable *by itself*: either by the signals emitted by itself, or by the foreknowable fact that it happens. It seems that the objections discussed above are no longer a threat to these two types of cases: we can barely say that the indeterministic processes are rendered deterministic by itself. In what follows, I will solely focus on the backward time travel case, since the foreknowledge case is more complicated and I think many issues have to be dealt with before we can go into it.²⁰

Now, let's examine the backward time travel case thoroughly. Recall that the objection against the actual blockage case is that the blockage mechanism might turn the indeterministic process into a deterministic one, by helping decide which neural pathway is going to be activated. However, I believe that the backward time travel case does not suffer from this objection. The main difference between the backward time travel case and the actual blockage case is that we do not have to set up the blockage mechanism *in advance*. We only have to let the indeterministic process flow naturally, and let the tachyonic signals be produced. We literally have to do *nothing* to

²⁰ Hunt mentioned the worry that foreknowledge may be an incoherent concept. (2003, 177-8)



the indeterministic process. Recall that in the actual blockage case, we have to set up the blockage mechanism before the indeterministic process P leads to the decision D. However, in the backward time travel case, the blockage mechanism comes *after* P has led to D, explanatorily, not temporally. The tachyonic signals indicating that P will lead to D can only be emitted after P has actually led to D. Thus, the tachyonic signals do not predetermine that P will lead to D; rather, P has *already* led to D indeterministically! P's actually leading to D comes explanatorily, also causally, prior to the tachyonic signals indicating that this will be the case. P's indeterministically leading to D causes the tachyonic signals being emitted, not the other way around. Hence, I do not find any reasons to say that the tachyonic signals have turned the indeterministic process into a deterministic one. Also, the fact that the tachyonic signals come explanatorily and causally after the indeterministic process supports the view that we should have the same intuition about moral responsibility before and after the signals are added to this scenario. If in the original indeterministic process, we think that Jones is morally responsible for killing Smith, then we should think the same after the tachyonic signals are added to this scenario. Since, the signals are not part of the actual causal sequence that contributes to the murder; they come *after* the actual causal sequence is finished. Given what has been said, I conclude that the backward time travel case escapes the objections leveled against the actual blockage case.

However, things are not that easy for the backward time travel case. For it to be a successful Frankase, it is not enough that it does not assume causal determination, and that it elicits the right intuition about moral responsibility. More importantly, it has to eliminate the alternative possibilities of the agent! It is this important feature, that the critics would claim that the backward time travel case fails to establish.

In the backward time travel case, Jones's killing Smith is an instance of agent-causation and indeterministic causation. However, due to the backward time travelling tachyonic signals, Hunt claimed that Jones's killing Smith is unavoidable. In one sense, the presence of the signals emitted from the future state in which Jones kills Smith does seem to render the murder unavoidable. Nevertheless, in another sense, claiming that an action, which is agent-caused unhinderedly by an indeterministic process, is unavoidable also seems to be mysterious in extreme, especially when the discussion is conducted under the context relevant to moral responsibility.

It would be helpful to restate the target that the backward time travel case is trying to falsify, PAP: a person is morally responsible for what he has done only if he could have done otherwise. Now, this question has to be answered: in the backward time travel case, *could* Jones do otherwise than killing Smith? I believe the answer can be found in David Lewis's discussion about the grandfather paradox.²¹ In discussing whether a time traveler can kill his own grandfather or not, Lewis had suggested that the word "can" is context-sensitive. That is to say, whether someone S can or cannot do something A should be decided under certain contexts. By "context", Lewis meant that we should first decide how inclusive a set of facts we are holding fixed. After that, we can see whether this set of facts is compossible with S's doing A. If they are compossible, then we say S *can* do A in a certain sense. Borrowing Lewis's example, consider whether Lewis can speak Finnish or not: if we hold fixed all the mental and physical features of Lewis, including the fact that he lacks any knowledge of Finnish, then of course the above set of facts is not compossible with Lewis's speaking Finnish. Thus, we would reach the conclusion that Lewis cannot speak Finnish in a more inclusive sense. However, if the discussion is conducted in a

²¹ See Lewis (1976).

physiology class, then we might only hold fixed the physiological properties of Lewis. In this case, the set of facts held fixed is compossible with Lewis's speaking Finnish. We can thus conclude that Lewis can speak Finnish in a physiological sense.

Back to the question we are now trying to answer: in the backward time travel case, *could* Jones do other than killing Smith? The answer would be: yes, and no. The context-sensitivity of the word "can" explains why we have conflicting intuitions on this issue: on one hand, due to the presence of the tachyonic signals, we think that Jones could not do otherwise; on the other hand, considering that the process leading to the murder is indeterministic and that the action is agent-caused, we also think that Jones could do otherwise. Now, let's say that in the backward time travel case, Jones's killing Smith happens at the time t_2 , and that the tachyonic signals the action emits travel back to, and are present at the earlier time t_1 . First, let's hold fixed the entire world states up to t_1 , *including* the presence of the tachyonic signals at t_1 . This set of facts is not compossible with Jones's doing otherwise than killing Smith at t_2 , since the presence of the signals at t_1 indicates that Jones kills Smith at t_2 (let's assume that Jones's killing Smith at t_2 is the only possible cause of the presence of the said signals at t_1). Given that it is impossible that Jones does not kill Smith at t_2 but also kills Smith at t_2 , we can conclude that Jones *could not* do otherwise than killing Smith at t_2 *in a more inclusive sense*, more inclusive comparing to the next sense. Second, this time let's hold fixed the entire world states up to t_1 , *except* the presence of the tachyonic signals at t_1 . This set of facts is compossible with Jones's doing otherwise than killing Smith at t_2 , since we are assuming indeterministic causation, nothing in the fixed set of facts stops Jones from not killing Smith. It is just that in the possible worlds in which Jones does not kill Smith at t_2 , some other tachyonic signals are present at t_1 . Hence, we can conclude that Jones *could* do otherwise than killing

Smith at t_2 *in a realistic sense* (I will explain why it is realistic later).

Now, the remaining question is this: which sense PAP is using, the more inclusive sense, or the realistic sense? Since, if we want to falsify PAP by the backward time travel case, we have to make sure that they are using the same sense.

We already know that the backward time travel case claims that Jones could not do otherwise than killing Smith at t_2 *in a more inclusive sense*. However, I think what PAP really means is this: a person is morally responsible for what he has done only if he could have done otherwise *in a realistic sense*. That is to say, the backward time travel case and PAP are not talking on the same basis. Hence, the backward time travel case fails to falsify PAP.

To see why the more inclusive sense is not the moral-responsibility-relevant sense, we have to make clearer what facts are held fixed in it. Notice that in this sense, we hold fixed the tachyonic signals at t_1 caused by Jones's killing Smith at t_2 . However, if Jones's killing Smith at t_2 is the only possible cause of the signals at t_1 , then in holding fixed the signals, we are actually also holding fixed the murder. Since, there are no close possible worlds in which the said signals are present at t_1 , but Jones does not kill Smith at t_2 .²² Hence, in holding fixed the entire world states up to t_1 , we are actually holding fixed those world states *plus* Jones's killing Smith at t_2 . Now, it is easier to see why the more inclusive sense is not the sense in which PAP makes its claim. An agent does some actions, and we, applying PAP, want to check whether he could do otherwise than those actions. Of course, in doing this, we should not hold fixed the fact that he actually does those action! When checking whether an agent could do otherwise than some A he actually does, we should always set aside the fact

²² By "close possible worlds," I mean the possible worlds which share the same physical laws (especially those concerning tachyons) with Jones and Smith's world.

that he actually does A, or else, the only answer we could get is that he could not do otherwise. Hence, when applying PAP, we should take a realistic context, under which meaningful discussions about alternative possibilities could be made. The context is realistic because it only focuses on factors that we care about alternative possibilities and moral responsibility; for example, the agent's physical and mental conditions, and the circumstance he is in (just to mention a few). We should only hold these facts fixed when discussing whether the agent could do otherwise.

In conclusion, I think that PAP claims that a person is morally responsible for what he has done only if he could have done otherwise *in a realistic sense*. However, the backward time travel case only shows that Jones could not do otherwise *in a more inclusive sense*. If we adopt the realistic sense, then what the backward time travel case would show is that Jones could actually do otherwise. Hence, the backward time travel case fails to escape from the “indeterministic horn” of the dilemma defense. Although the actual causal sequence in it is an indeterministic process, nevertheless the agent actually could do otherwise.

To sum up, in this section, I have critically discussed two types of blockage cases: the actual blockage case, and the backward time travel case. I conclude that they both fail to escape from the dilemma defense. I believe the failure of the blockage cases does not show their doom; it rather points to a new path that should be taken. In what follows, I will try to develop a new version of the blockage case.

Chapter 3



A New Type of Blockage Case

3.1 The Illuminated Way

Now, it is time to take stock. What do all the above discussions suggest? To build a successful Frankase, we must find a way to respond to the dilemma defense. First, we must avoid assuming causal determination in the actual causal sequence. The actual causal sequence leading to the target action (the action that the agent is morally responsible for) has to be an indeterministic process. That is to say, given all the states of this process before the target action takes place, no action is determined to happen. Second, we must manage to eliminate the alternative possibilities of the agent, at least those that are crucial to PAP. In addition, we must accomplish this *outside* the actual causal sequence. That is to say, we have to eliminate the alternative possibilities in a way that, viewed within the actual causal sequence, no action is determined to happen; however, if we take our mechanism into account, the action actually performed is the only possible action for the agent. Last but not least, we must be able to argue that the agent is morally responsible for the target action. Difficult as it seems, at least we already know what requirements we have to meet in order to build a successful Frankase.

3.2 Developing a New Type of Blockage Case



I believe that the motivation behind the proponents of the Frankases is clear: we believe that the *control* pertinent to free will and moral responsibility can only be manifested by what happens in the *actual causal sequence*, not by the possession of the ability to do otherwise. That is to say, the control in what actually happens, is all the control that is needed for free will and moral responsibility.

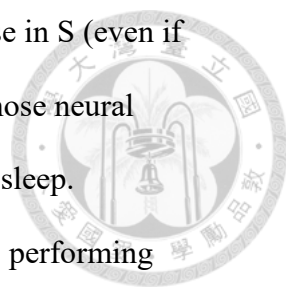
Furthermore, I also believe that PAP is a principle that needs clarification. Although I think PAP is false, it is still important to make clear what the proponents of PAP have in mind. Recall that PAP only says that a person is morally responsible for what he has done only if he could have done otherwise. However, the “only if” relation only states that the ability to do otherwise is a necessary condition for moral responsibility, but that is too vague. There are different levels of necessary conditions. For example, being alive is also a necessary condition for a person to be morally responsible for his actions, but it is evident that being alive and having the ability to do otherwise are not on the same level. The difference between them is that having the ability to do otherwise is a *constitutive* or *explanatory* condition for moral responsibility (for the proponents of PAP), while being alive is not. The proponents of PAP should maintain that the ability to do otherwise is *essential* to moral responsibility, and that having the ability to do otherwise partially *explains* why the agent is morally responsible for his actions. The ability to do otherwise is not just any necessary condition for moral responsibility, but an essential one. Hence, I think PAP should be more detailed. I believe what the proponents of PAP actually have in mind is this: a person is morally responsible for what he has done only if, *and in that case partially because*, he could have done otherwise. Call this PAP+. PAP+ is a more

detailed version of PAP, and I believe that it is what the proponents of PAP is actually defending.²³ PAP is a weaker principle than PAP+, since it only states the necessary condition. I believe that hiding behind a weaker principle while actually claiming a stronger principle is unfair to their opponents, since it makes them harder to construct a counterexample.

Given what has been said, we can now start to construct a new blockage case. First, we need an indeterministic actual causal sequence by which no action is determined to be brought about, to avoid begging the question against the incompatibilists about moral responsibility. Second, we need to eliminate the alternative possibilities that allegedly explain the agent's moral responsibility, but we have to use a mechanism that does not engage in the actual causal sequence. Finally, we have to argue that the agent is morally responsible for his actions. The following is my blockage case, Brain Transformation.

Tim is an ordinary person. One night, an evil neural scientist, Judy, sneaks into his house. Judy wants to test her newly-developed brain transformation technology on Tim. While Tim is asleep, Judy performs a surgery on him, and Tim is totally unaware of the whole thing. Judy transforms Tim's brain. After the surgery, the functions of Tim's brain have been changed. Until the end of his life, only one specific sequence of neural activities remains the same function as before. Call this sequence of neural activities S. If Tim's brain performs the exactly same neural activities as those in S, then it would function normally, doing the jobs that those neural activities originally do. However, the functions of every sequence of neural activities other than S have been changed. They only do one job now: making Tim fall into permanent sleep. That

²³ Fischer (1999) shared a similar view. He argued that the control that the proponents of PAP are looking for cannot be provided by the mere existence of any sorts of alternative possibilities, but can only be provided by the "robust alternatives."



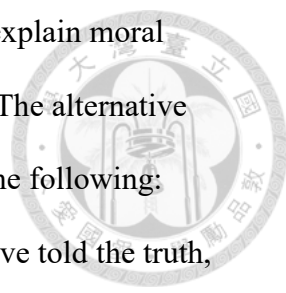
is, if Tim's brain performs some neural activities different from those in S (even if they are only different in order), then it would not do the jobs that those neural activities originally do; it would only make Tim fall into permanent sleep. Metaphorically speaking, Tim's brain has only two ways to go now: performing exactly the same neural activities as those in S, or falling into permanent sleep. Tim has no idea that his brain has been transformed in this way. However, after the surgery, Tim does not fall into permanent sleep. Actually, he lives his life normally, and his brain works well. Nothing strange happens until his death. It turns out that, due to sheer luck, Tim's brain follows S for his whole life! After the surgery, Tim's brain has been performing exactly those neural activities of S, in the exact order of S, hence it functions just as before. Tim has a moral life like many of us: he deliberates whether to do good deeds or bad deeds, and has done both throughout his life. Furthermore, and most importantly, all the big decisions in his life are indeterministically caused, in the sense that any earlier states of Tim do not determine any of his later decisions. Hence, given his states up to the moment before the decision is made, Tim could either make the decision that he actually makes, or perform other neural activities whose function is making Tim fall into permanent sleep. Thus, no causal determination is assumed in his decision-making processes.

I believe that this is a new type of blockage case that avoids the objections faced by the actual blockage case. Let's examine the actual causal sequences leading to Tim's decisions. Note that Tim's decision is *singly caused* by his deliberation process, unlike in the actual blockage case, Jones's decision is *jointly caused* by his deliberation process and the blockage mechanism. This is because, the brain transformation does not block any neural activities. It only changes the functions of them. That is to say, it is totally open for Tim to perform other neural activities. The

brain transformation does not in any way restrict the performance of any neural activities. This proves that when the neural activities of S are actually performed by Tim, the brain transformation has contributed nothing to this fact, since it puts no restriction on which neural activities would be performed. We can be sure that even if Tim's brain has not been transformed, his whole life would still be the same. Hence, I conclude that all of Tim's decisions are *singly caused* by his deliberation processes, and that the brain transformation is *not* part of the actual causal sequences that lead to these decisions.

Also, I claim that Tim is morally responsible for those (moral) decisions he makes in his life. My case does not beg the question against the incompatibilists about moral responsibility, since there is no causal determination in the actual causal sequences leading to the decisions, as I have explained above. The brain transformation does not affect Jones's status as a moral agent. Since Tim's brain follows S for his whole life, it functions properly. He has not lost any ability that is required to be a moral agent. Moreover, he has not been coerced or manipulated to make decisions. He is as accountable as any normal person.

Finally, and most importantly, I believe that Brain Transformation is a counterexample to PAP+. PAP+ says that a person is morally responsible for what he has done only if, *and in that case partially because*, he could have done otherwise. In Brain Transformation, Tim is morally responsible for what he has done. Also, Tim could have done otherwise, since he could perform other neural activities. However, the crux of Brain Transformation is that Tim's moral responsibility is *not explained by* the fact that he could have done otherwise. Let me elaborate. Apart from those functions that Tim's brain has actually performed throughout his life, there is only one kind of function that it could have performed: to make Tim fall into permanent sleep.



However, this kind of alternative possibility is not the one that can explain moral responsibility, or can be said to be essential to moral responsibility. The alternative possibilities that the proponents of PAP+ regard as crucial are like the following: Jones is morally responsible for lying to Smith, because he could have told the truth, or have said something irrelevant, or even have said nothing. Suppose at this moment, Jones could also have performed a special neural activity whose function is to make him fall into permanent sleep, hence avoiding lying. It would be ridiculous for the proponents of PAP+ to say that Jones is morally responsible for lying to Smith, (partially) because he could have performed the neural activity that make him fall into permanent sleep. This kind of alternative possibility is totally irrelevant to the moral responsibility under discussion. If the proponents of PAP+ wish to explore the nature of moral responsibility, and explain moral responsibility, then this kind of alternative possibility is not what they are looking for. However, this kind of alternative possibility is the only alternative possibilities present in Brain Transformation. All other alternative possibilities that allegedly explain Tim's moral responsibility are eliminated. Hence, we seem to have a scenario in which the agent is morally responsible for his actions, nevertheless lacks all the alternative possibilities that the proponents of PAP+ regard as necessary to moral responsibility. Brain Transformation is a counterexample to PAP+.

3.3 PAP or PAP+?

As one can notice, Brain Transformation is a counterexample to PAP+, not PAP. Hence, if there is a group of proponents of PAP who only support PAP, but do not believe in PAP+, then they would not be affected by Brain Transformation. However,

I believe that the main proponents of PAP (the leeway theorists) also support PAP+, and that for them, PAP without PAP+ is a hollow principle that is very hard to be defended. In what follows, I will examine the arguments of two significant proponents of PAP, Widerker and David Copp. In examining their arguments for PAP, I wish to bring out their implicit commitment to PAP+. However, there may actually be a kind of source theorists who only support PAP, but refuse PAP+. I will argue that their position is amiable to mine.

In his (2003), Widerker developed his famous “W-defense.” Widerker claimed that even if IRR-circumstances are possible, agents in them are not morally responsible for their actions:

Let me grant, for the sake of discussion, that in the IRR-situation under consideration, Jones acted freely in the sense that what he did he did for reasons of his own without being causally determined or coerced to so act. Still, since you, Frankfurt, wish to hold him blameworthy for his decision to break his promise, tell me *what, in your opinion, should he have done instead?*²⁴ Now, you cannot claim that he should not have decided to break the promise, since this was something that was not in Jones’s power to do. Hence, I do not see how you can hold Jones blameworthy for his decision to break the promise. (63)

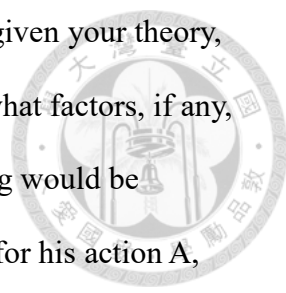
Following this line of thought, Widerker proposed PAE: an agent S is morally blameworthy for doing A only if under the circumstances it would be morally reasonable to expect S not to have done A. (63) Widerker defended PAP by adding an

²⁴ Widerker and Copp only focused on one side of moral responsibility, that is, moral blameworthiness.

extra constraint on ascriptions of moral blame; that is, only when it is morally reasonable to name the alternative actions that the agent should have done, would he be morally blameworthy for his action. The need of this extra constraint implies that in the minds of the proponents of PAP, there are some thoughts that cannot be articulated solely by PAP.

Copp is another defender of PAP. In his (2003), he claimed that PAP can be derived from the intuitive maxim, “ought” implies “can,” which he called “the Maxim.” Copp argued that the Maxim is well-motivated. He provided two reasons to accept the Maxim. Given that PAP can be derived from the Maxim, Copp believed that the two reasons supporting the Maxim can also support PAP. The first reason supporting the Maxim is that it is unfair to require someone to do something, if he were unable to do it. (271-2) Applying this to PAP, Copp claimed that if a person cannot act otherwise, then it would be unfair to morally require him to act otherwise. Thus, the person is not morally blameworthy for the action that he actually does. The second reason is that morality is meant to guide our decision-making. (273-4) That is to say, if we are morally required to do some action A in a situation, then it means that all options other than A are morally ruled out in that situation. However, if the morally required action is something that we cannot do, then all of our options are ruled out. If this were to be the case, then morality could not guide our decision-making in this situation. Hence, the morally required action has to be an action that the agent can perform. Applying this to PAP, Copp maintained that if morality is to be decision-guiding, then an agent could not be morally required not to do an action A when not doing A is not one of his genuine options. Thus, his actually doing A would not be morally blameworthy.

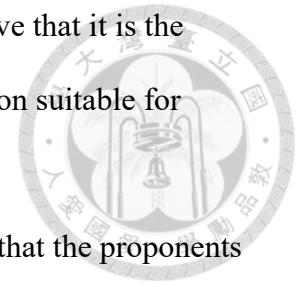
I believe that Widerker and Copp are representative theorists that defend PAP by



moral arguments. Now, we can ask them an important question: if, given your theory, a person is morally blameworthy for his action in a situation, then what factors, if any, *explain* his moral blameworthiness? The answers that I am expecting would be something like this: given PAP, if a person is morally blameworthy for his action A, then it must be the case that he could do otherwise than A, since in this way we could name the alternative actions that he should do, and be morally reasonable to expect him not to do A; also, in this way it would be fair to require him not to do A, and that morality could tell him what should he do instead; hence, to answer your question, I believe that *the fact that he could do otherwise than A at least partially explains the fact that he is morally blameworthy for doing A*; still, there may be other partial explanations for his moral blameworthiness.

It seems to me that, from Widerker and Copp's arguments for PAP, we can see that they are conveying more thoughts than those that are contained in PAP. The alternative possibilities are not only a necessary condition for moral responsibility which, in every possible world, agents who are morally responsible for their actions have alternative possibilities available to them; moreover, the accessibility to alternative possibilities is exactly *why* the agents are morally responsible for their actions. According to Widerker and Copp's arguments, a person is morally responsible for his action, only if it is morally reasonable for us to point out the alternatives that should be taken, and that it is fair and decision-guiding to hold him morally responsible. Now, it seems that they are not only arguing for the *necessity* of alternative possibilities, they are also arguing for the *role* played by alternative possibilities in ascriptions of moral responsibility, which is to make ascriptions of moral responsibility reasonable, fair, and decision-guiding. Hence, for them, alternative possibilities are not simply a necessary condition, but a factor that has

substantial effects in ascriptions of moral responsibility. They believe that it is the accessibility to alternative possibilities that (partially) makes a person suitable for being held morally responsible for his action.



With this much being said, I believe that it is safe to conclude that the proponents of PAP who share similar views with Widerker and Copp also support PAP+. That is to say, when asked what *explains* an agent's moral responsibility for his action, one of their answers (also an important one) would be: the accessibility to alternative possibilities. It would be strange if, with the central role played by alternative possibilities in the ascriptions of moral responsibility, the former just cannot explain the latter. As long as they support PAP+, Brain Transformation would be a counterexample to their view.

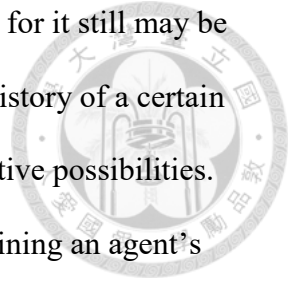
To be more careful, are there any reasonable kinds of genuine leeway view that only support PAP, but do not need the help of PAP+? I do not think so. The reason is that if a leeway theorist refuses that the availability of alternative possibilities explains an agent's moral responsibility, then I do not see how he could defend the view that the former is necessary for the latter within the framework of the leeway view. Perhaps he could argue that there are some *other factors* that explain an agent's moral responsibility, and that the availability of alternative possibilities is necessary for those factors. However, in doing so, he is actually moving away from the leeway view, since he believes that the ground of moral responsibility is something different from alternative possibilities. On the other hand, if he does not invoke other factors, and tries to defend PAP directly, then he would have to explain the special connection between alternative possibilities and moral responsibility. The more he says about this special connection, the closer he would get to PAP+; just like what Widerker and Copp had done. In sum, I believe it is hard, if not impossible, to defend PAP without

the help of PAP+ within the framework of the leeway theory.

However, outside the framework of the leeway theory, there is indeed one kind of position, which may or may not support PAP, but if it does, it does not need to support PAP+; that is the *source incompatibilism*. Source incompatibilists believe that what explains the ascriptions of moral responsibility is not the accessibility to alternative possibilities, but being the source or origination of the action. Nonetheless, they also believe that determinism is incompatible with moral responsibility, since determinism makes our actions determined by factors external to our control. Hence, if determinism is true, then the sources of our actions would not be us. As had been pointed out in Michael Della Rocca (1998), the above type of incompatibilism may consider alternative possibilities as necessary for moral responsibility. However, alternative possibilities are needed only to establish indeterminism, not to explain the ascriptions of moral responsibility. Although the source incompatibilists believe that the ground of moral responsibility is being the source of one's actions, they may nevertheless think that having genuine alternative possibilities is necessary for indeterminism, and that indeterminism is necessary for being the source of one's actions. Hence, they may support PAP, since the availability of alternative possibilities is necessary for the ascriptions of moral responsibility; but they do not support PAP+, since the former does not explain the latter. Pereboom is a prominent example of the source incompatibilists. In his (2003), following his own "necessary-condition Frankase," he articulated exactly the same thought:

This Frankfurt-style argument indicates that the availability of alternative possibilities is not relevant *per se* to explaining an agent's moral responsibility for an action. It is compatible with this conclusion that the availability of

alternative possibilities is necessary for moral responsibility, for it still may be that moral responsibility requires an indeterministic causal history of a certain sort, and that such a history entails the availability of alternative possibilities. But these alternative possibilities would be relevant to explaining an agent's moral responsibility not *qua* alternative possibilities, but only because they indicate a causal history of a kind that is relevant *per se* to explaining an agent's moral responsibility. (197)



What the above discussion shows is that Brain Transformation is not aiming at the source incompatibilists. I build up Brain Transformation only to argue against those who believe that the availability of alternative possibilities can *explain* an agent's moral responsibility. To repeat, Brain Transformation is a counterexample to PAP+, not to PAP. Hence, if some source incompatibilists wish to accept PAP, they are welcomed to do so. After all, as long as they believe that the ground of moral responsibility is being the source of one's actions, not the availability of alternative possibilities, then the source incompatibilists and I are on the same page.

In conclusion, in this section I claim that all genuine leeway theorists support PAP+, and that the only position supporting PAP but not PAP+ is the source incompatibilism.

3.4 Conclusion

In this chapter, I develop a new type of blockage case, Brain Transformation. I believe Brain Transformation shows that the availability of alternative possibilities *cannot* explain an agent's moral responsibility. Furthermore, I argue that all genuine

leeway theorists support PAP+, which has been falsified by Brain Transformation.

Hence, I think that Brain Transformation is a threat to the leeway theorists.



Chapter 4



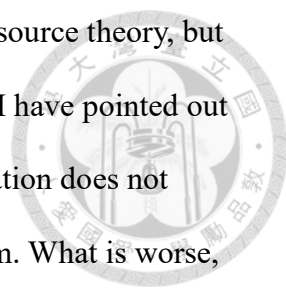
Toward the Source Compatibilism

4.1 Preliminaries

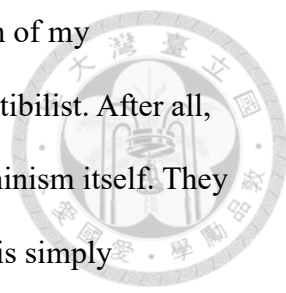
As I have mentioned in Chapter 1, the aim of this thesis is to provide a defense for the source compatibilism. This job has not been done yet. At best, Brain Transformation can only show that the availability of alternative possibilities is irrelevant to the explanation of an agent's moral responsibility. At this point, there are still lots of works to do.

I would like to start with a relatively easy task. In Brain Transformation, if Tim's only alternative possibility cannot explain his moral responsibility for his moral choices, then what does? I believe that the best candidate here is the fact that Tim makes those choices *by himself*. After deleting all the alternative possibilities in this scenario that seem to explain his moral responsibility, this point comes out clearly, since without the alternative possibilities, the only factor left that can plausibly serve as the ground of moral responsibility is the *sourcehood* of the actions. That is to say, Tim is morally responsible for his choices because he is the source of those choices. No one but he produces those choices, hence he should take responsibility for them.

Hence, the lesson taught by Brain Transformation is that the availability of alternative possibilities is irrelevant to the explanation of an agent's moral responsibility, and that what is relevant is the sourcehood of the actions. Now, with



this lesson, Brain Transformation is a good example supporting the source theory, but the compatibilist part of the theory has not yet been established. As I have pointed out in 3.3, source incompatibilism is still on the table. Brain Transformation does not show that Tim's moral responsibility is compatible with determinism. What is worse, the setting of Brain Transformation is an *indeterministic* world. This may give my readers an impression that I think that moral responsibility is only possible in an indeterministic world, but this is not so. I set Brain Transformation in an indeterministic world only for the sake of argument. I make Tim's deliberation process indeterministic only because I want to avoid begging the question against the incompatibilists about moral responsibility. The indeterministic part of Brain Transformation is totally dispensable for me. We can imagine a deterministic version of Brain Transformation: The surgery that Tim has undergone is the same. The only difference is that Tim lives in a world where determinism is true. Hence, although it is physically possible that he performs some other neural activities, it is determined by prior world states and natural laws that he will not. In other words, it is determined that Tim will perform those neural activities that he actually performs throughout his life. In this version of Brain Transformation, is Tim still morally responsible for the moral choices he makes in his life? My intuition remains the same. I believe that Tim is still morally responsible for those choices. My reasoning is that even if determinism is true, the brain transformation surgery still has no causal impact on Tim's moral choices. The surgery merely changes the functions of some neural activities; it says nothing about which neural activities are going to be performed. Hence, I think we can say that Tim lives his life "as if the surgery has never happened." All of the choices are made by him (though deterministically), thus he should take responsibility for them.



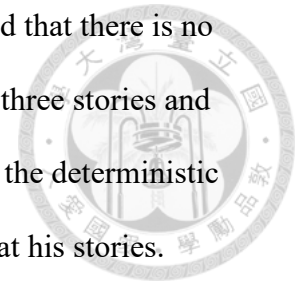
However, the above reasoning is little more than an articulation of my compatibilist intuition. I surely do not expect it to move an incompatibilist. After all, what the incompatibilists are worrying is not the surgery, but determinism itself. They believe that in the deterministic version of Brain Transformation, it is simply impossible for Tim to be morally responsible for his choices, no matter whether Tim has undergone the surgery. Hence, to defend source compatibilism, Brain Transformation is not enough. Even in its deterministic version, all that I can do is to repeat that “according to the compatibilist theories,” Tim is morally responsible for his choices. This is not going to persuade my rivals. To defend source compatibilism, I have to respond to the criticisms launched against my position by the incompatibilists. Hence, in this chapter, I will review the strongest attack against compatibilism, the manipulation arguments, and respond to it.

The manipulation arguments have two famous examples, Pereboom’s four-case argument (2001, 2014) and Mele’s zygote argument (2006, 2019). I will examine both, argue that the former had been successfully answered by Mele (2006), and respond to the latter.

4.2 Pereboom’s Four-case Argument

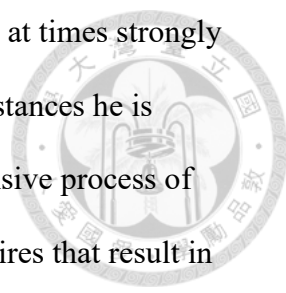
In Pereboom (2001) and (2014), Pereboom proposed his four-case argument. The aim of the four-case argument is to argue that even if the source theory is true (as Pereboom himself believes), moral responsibility is still incompatible with determinism. In this argument, Pereboom spun four similar stories. In the first three stories, the agents are somehow manipulated. In the last one, the agent simply lives in a deterministic world. Pereboom tried to elicit the intuition that the agents in the first

three stories are not morally responsible for their actions, and argued that there is no relevant difference regarding moral responsibility between the first three stories and the last one. Hence, he concluded that we should judge the agent in the deterministic world as not morally responsible for his actions. Let us take a look at his stories.



Case 1: A team of neuroscientists has the ability to manipulate Plum's neural states at any time by radio-like technology. In this particular case, they do so by pressing a button just before he begins to reason about his situation, which they know will produce in him a neural state that realizes a strongly egoistic reasoning process, which the neuroscientists know will deterministically result in his decision to kill White. Plum would not have killed White had the neuroscientists not intervened, since his reasoning would then not have been sufficiently egoistic to produce this decision. But at the same time, Plum's effective first-order desire to kill White conforms to his second-order desires. In addition, his process of deliberation from which the decision results is reasons-responsive; in particular, this type of process would have resulted in Plum's refraining from deciding to kill White in certain situations in which his reasons were different. His reasoning is consistent with his character because it is frequently egoistic and sometimes strongly so. Still, it is not in general exclusively egoistic, because he sometimes successfully regulates his behavior by moral reasons, especially when the egoistic reasons are relatively weak. Plum is also not constrained to act as he does, for he does not act because of an irresistible desire – the neuroscientists do not induce a desire of this sort.

Case 2: Plum is just like an ordinary human being, except that a team of neuroscientists programmed him at the beginning of his life so that his



reasoning is often but not always egoistic (as in Case 1), and at times strongly so, with the intended consequence that in his current circumstances he is causally determined to engage in the egoistic reasons-responsive process of deliberation and to have the set of first and second-order desires that result in his decision to kill White. Plum has the general ability to regulate his actions by moral reasons, but in his circumstances, due to the strongly egoistic nature of his deliberative reasoning, he is causally determined to make his decision to kill. Yet he does not decide as he does because of an irresistible desire. The neural realization of his reasoning process and of his decision is exactly the same as it is in Case 1 (although their causal histories are different).

Case 3: Plum is an ordinary human being, except that the training practices of his community causally determined the nature of his deliberative reasoning processes so that they are frequently but not exclusively rationally egoistic (the resulting nature of his deliberative reasoning processes are exactly as they are in Cases 1 and 2). This training was completed before he developed the ability to prevent or alter these practices. Due to the aspect of his character produced by this training, in his present circumstances he is causally determined to engage in the strongly egoistic reasons-responsive process of deliberation and to have the first and second-order desires that issue in his decision to kill White. While Plum does have the general ability to regulate his behavior by moral reasons, in virtue of this aspect of his character and his circumstances he is causally determined to make his immoral decision, although he does not decide as he does due to an irresistible desire. The neural realization of his deliberative reasoning process and of the decision is just as it is in Cases 1 and 2.

Case 4: Everything that happens in our universe is causally determined by virtue of its past states together with the laws of nature. Plum is an ordinary human being, raised in normal circumstances, and again his reasoning processes are frequently but not exclusively egoistic, and sometimes strongly so (as in Cases 1–3). His decision to kill White issues from his strongly egoistic but reasons-responsive process of deliberation, and he has the specified first and second-order desires. The neural realization of Plum’s reasoning process and decision is exactly as it is in Cases 1–3; he has the general ability to grasp, apply, and regulate his actions by moral reasons, and it is not because of an irresistible desire that he decides to kill. (2014, 76-9)

Notice that in all four cases, determinism obtains. Pereboom pointed out that in Case 1, we have a clear intuition that Plum is not morally responsible for killing White, since the murder seems to be the result of the neuroscientists’ intensions. However, Plum satisfies all the compatibilist conditions for an agent to be morally responsible for his actions. Hence, only by Case 1, Pereboom thought that we can already establish that all the allegedly sufficient compatibilist conditions fail. (2014,77) Moving on to Case 2, Pereboom believed that we have the same intuition that Plum is not morally responsible for killing White. In Case 3, if a compatibilist wants to argue that Plum is morally responsible for killing White, he would have to point out the difference between Case 3 and the former two cases that justifies his different judgments in moral responsibility. This, Pereboom claimed, cannot be achieved, since in all three cases, Plum meets all the compatibilist conditions for morally responsible actions. (2014, 78). The situation is the same in Case 4. The compatibilists cannot point out the relevant difference between it and the first three

cases regarding moral responsibility. Hence, the compatibilists are forced to conclude that even in a normal scenario in which determinism is true, Plum is not morally responsible for killing White. Even if they have strong intuition that Plum is morally responsible, due to the force of this argument, they have to revise it. Furthermore, Pereboom claimed that the best explanation of our non-responsibility judgments in the four cases is that Plum's actions "are causally determined by factors beyond his control." (2014, 79) In conclusion, the four-case argument shows that in a deterministic world, moral responsibility is impossible because all of our actions are causally determined by factors beyond our control.

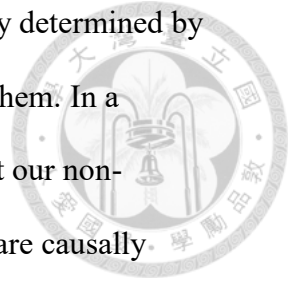
In Mele (2006), Mele offered a powerful reply to the four-case argument, which I believe is successful.²⁵ Mele pointed out that actually, the indeterministic version of the first three cases can elicit the same non-responsibility intuition. Imagine that in Case 1, the neuroscientists manipulate Plum "by means of an indeterministic mechanism. There is an extremely good chance that each push of a button will have the result the neuroscientists want, but each push also has a tiny chance of incapacitating Plum. These are the only possible outcomes of the button pushes. As it happens, Plum is not incapacitated." (2006, 140) Mele thought that our intuition would remain the same. I agree. If in the original case we have the intuition that Plum is not morally responsible for killing White, then adding a tiny chance of failing to manipulate Plum would not change our verdict. After all, Plum is still severely manipulated. We can do the same thing to Case 2 and 3, and get the same non-responsibility intuitions in them (if you do have ones in the original cases). Mele argued that, if we have the same intuitions in both deterministic and indeterministic

²⁵ Mele targeted at the 2001 version of the four-case argument, but the 2014 version is almost the same as the 2001 one, hence Mele's objection is still valid against the 2014 version.

versions of Case 1-3, then this shows that the determinism in these cases are dispensable. That is to say, determinism is not the factor that elicit our non-responsibility intuitions. Now, the common factor left in the cases and their analogues is the fact that *Plum is manipulated*. This fact best explains why we believe that Plum is not morally responsible for killing White, no matter the neuroscientists manipulate him by deterministic or indeterministic mechanisms. Hence, the best explanation of our intuitions in Case 1-3 is not that Plum's actions are causally determined by factors beyond his control, as Pereboom contended, but the fact that Plum is manipulated to kill White. This breaks the link between the first three cases and Case 4. In Case 4, there is no manipulation that supports our non-responsibility intuitions in the first three cases, hence, there *is* a relevant difference justifying asymmetric judgments on moral responsibility. Hence, the compatibilists are not forced to say that Plum is not morally responsible for killing White in a normal deterministic scenario. (2006, 139-44)

It would be helpful to clarify the dialectic here. In Case 1-3, there are two different elements in a scenario, the determinism element (the fact that causal determinism obtains) and the manipulation element (the fact that Plum is manipulated to kill White). Which element is doing the job of explaining our non-responsibility intuitions? Actually, for the four-case argument to succeed, it cannot be the case that the manipulation element is doing this job. Since, as Mele (2006) pointed out, there is no manipulation in Case 4, hence asymmetric judgments between Case 4 and Case 1-3 on moral responsibility would not be unwarranted. That is to say, proponents of the four-case argument have to argue that it is actually the determinism element that is inducing our non-responsibility intuitions in Case 1-3, and that the manipulation element is dispensable. The job of the manipulation element is only to help bring out

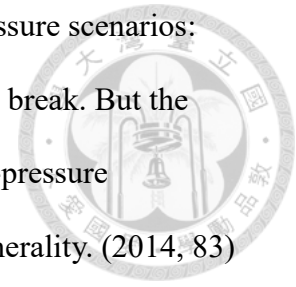
the intuition that we tend to ignore; that is, if our actions are causally determined by factors beyond his control, then we are not morally responsible for them. In a manipulation-free, normal deterministic scenario, we tend to neglect our non-responsibility intuition because we are blind about how our actions are causally determined by factors beyond our control. The manipulation element helps present this fact to us vividly.



Pereboom replied to Mele’s objection to the four-case argument in Pereboom (2014). He claimed that even if we have the same non-responsibility intuitions in the indeterministic analogues of Case 1-3, this still does not show that causal determinism is dispensable in Case 1-3. He pointed out that when factors A and B can both give rise to a phenomenon P, it is possible that A and B explain P respectively, and that there is a more general explanation lying behind A and B that explains P. Consider his analogy:

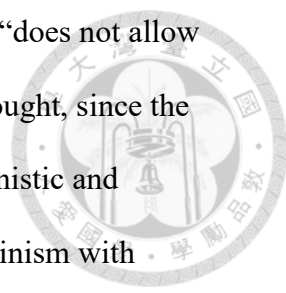
Imagine that a dam at one end of a reservoir would break if the reservoir were filled with more than one billion gallons of water, because the dam could not withstand the pressure that this volume of water would exert. Suppose the reservoir is in fact filled with more than one billion gallons of water, and the dam breaks. It is natural to say here: “what explains the dam’s breaking is the water pressure.” However, someone might object: “if the reservoir were filled with more than one billion gallons of oil, it would also have broken. So the water pressure doesn’t explain the dam’s breaking.” To this the correct response would be: some true causal explanations set out the actual sufficient conditions for an event’s occurring, and accordingly the explanation by way of the water pressure is true. But there is also an explanation of the dam’s

breaking common to both the water pressure and the oil pressure scenarios: liquid pressure higher than a certain level caused the dam to break. But the water-pressure explanation doesn't compete with the liquid-pressure explanation—they are explanations at different levels of generality. (2014, 83)



Applying this to the four-case argument, Pereboom claimed that the determinism and indeterminism in Case 1-3 and their analogues both explain our non-responsibility intuitions. In deterministic cases, we have non-responsibility intuitions because Plum's actions are causally determined by factors beyond his control. In indeterministic cases, our non-responsibility intuitions stem from the fact that indeterminism "does not allow for the agent to settle whether the decision will occur." (2014, 84) Behind these explanations lies the more general explanation: the agents act under "causal circumstances that preclude responsibility-relevant control." (84) Determinism and indeterminism both achieve this, through different ways. Hence, Pereboom concluded that Mele's indeterministic analogues do not undermine the power of the four-case argument. In Case 1-3, the best explanation of our non-responsibility intuitions is still the fact that Plum's actions are causally determined by factors beyond his control.

I do not think that Pereboom's reply is successful. I think his diagnosis of Mele's indeterministic analogues is not entirely correct. Recall the indeterministic analogue of Case 1: the neuroscientists manipulate Plum "by means of an indeterministic mechanism. There is an extremely good chance that each push of a button will have the result the neuroscientists want, but each push also has a tiny chance of incapacitating Plum. These are the only possible outcomes of the button pushes. As it happens, Plum is not incapacitated." (Mele 2006, 140) I believe that in this scenario,



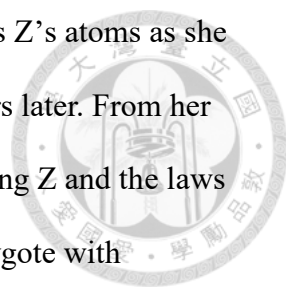
we have the non-responsibility intuition not because indeterminism “does not allow for Plum to settle whether the decision will occur,” as Pereboom thought, since the chance that Plum is incapacitated is so tiny! Comparing the deterministic and indeterministic versions of Case 1, we can see that replacing determinism with indeterminism actually does not make a big difference to our intuition, since the indeterminism introduced is minute. Whatever reason we have for having the non-responsibility intuition in the deterministic Case 1, we can expect that the same reason would explain our non-responsibility intuition in the indeterministic Case 1 (if we do have such intuition). If our intuition remains the same, it would be strange that adding a tiny chance of Plum being incapacitated changes the reason backing up our one and the same intuition. As Mele nicely pointed out in his (2019), “If Agnes (readers) were asked to identify a common feature of the two stories that supports her nonresponsibility judgments, she might say that, in both stories, Ernie (Plum) is Diana’s (the neuroscientists’) tool for getting things done that she (they) wants to have done. In my deterministic story, she might say, Ernie (Plum) is a perfect tool whereas in the indeterministic story he is a defective tool—a tool that might break down. She may view both kinds of toolhood as sufficient for Ernie’s (Plum’s) nonresponsibility for his actions.” (104, fn.2) Although Mele was talking about his zygote argument, the same explanation also goes for the four-case argument. I think that in both versions of Case 1, we have the non-responsibility intuitions because we view Plum as the neuroscientists’ tool. Plum is “used” by the neuroscientists to realize their intentions. Hence, I believe that Pereboom’s diagnosis of the indeterministic version of Case 1 is wrong. In indeterministic Case 1, indeterminism by itself does not preclude responsibility-relevant control, since the chance that Plum’s reasons cannot effectively lead to actions (in this case, that he is incapacitated) is so tiny. Imagine an

indeterministic version of Case 4: Plum is a normal human being; the only difference is that he always has a tiny chance of breaking down. Suppose that he never breaks down in his whole life. It is not so straightforward that this level of indeterminism would preclude his responsibility-relevant control. Our intuition would not change simply because this level of indeterminism is added.

To summarize, I believe that the best explanations of our non-responsibility intuitions in the deterministic and indeterministic versions of Case 1-3 are the same: the fact that Plum is manipulated to kill White. That is to say, Mele's indeterministic analogues successfully show that it is the manipulation element that elicits our non-responsibility intuitions in Case 1-3, not the determinism element. Hence, inference from Case 1-3 to Case 4 is unwarranted. The four-case argument does not succeed.

4.3 Mele's Zygote Argument

Actually, the manipulation arguments have another way to go. Recall that in the last section, Mele has spelled out that what is doing the intuition-inducing work in the manipulation arguments is the manipulation element, not the determinism element. In fact, this finding can be embraced by the proponents of the manipulation arguments. Mele acknowledged this, but still thought that a simpler form of manipulation argument has its power. Even though it is the manipulation element that is eliciting our non-responsibility intuition, with a carefully designed scenario, it is still possible to argue that there is no significant difference concerning moral responsibility between the manipulation scenario and the normal causal determinism scenario. Mele's zygote argument is one great example:

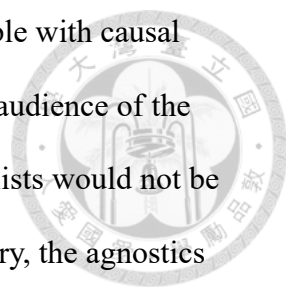


Diana (a Goddess) creates a zygote Z in Mary. She combines Z's atoms as she does because she wants a certain event E to occur thirty years later. From her knowledge of the state of the universe just prior to her creating Z and the laws of nature of her deterministic universe, she deduces that a zygote with precisely Z's constitution located in Mary will develop into an ideally self-controlled agent who, in thirty years, will judge, on the basis of rational deliberation, that it is best to A and will A on the basis of that judgment, thereby bringing about E.... Thirty years later, Ernie is a mentally healthy, ideally self-controlled person who regularly exercises his powers of self-control and has no relevant compelled or coercively produced attitudes. Furthermore, his beliefs are conducive to informed deliberation about all matters that concern him, and he is a reliable deliberator. (2006, 188)

From this story, Mele formulated his argument, ZAM:

1. Ernie is not morally responsible for anything he does.
2. Concerning moral responsibility of the beings into whom the zygotes develop, there is no significant difference between the way Ernie's zygote comes to exist and the way any normal human zygote comes to exist in a deterministic world.
3. So in no possible deterministic world in which a human being develops from a normal human zygote is that human being morally responsible for anything he or she does. (2013, 176)

According to Mele, premise 1 is supported by the intuition of the agnostics about



the question whether free will and moral responsibility are compatible with causal determinism. The agnostics are whom Mele considered as the right audience of the zygote argument, since it is expected that the “seasoned” compatibilists would not be moved by this argument.²⁶ Mele believed that upon hearing this story, the agnostics would sincerely report that Ernie is not morally responsible for anything he does. Mele claimed that their intuition is caused by a fact *F*, “the fact that just by assembling the atoms of Ernie’s zygote as she does and implanting them in Mary when she does, Diana intentionally brings it about and intentionally ensures that Ernie will do $A_1 \dots A_n$, where that is a complete, ordered list of Ernie’s actions.” (2019, 102) Now it is very clear that Mele embraced the manipulation element. *F* is the manipulation element in Ernie’s story. Mele was explicit that what elicits the agnostics’ non-responsibility intuition is the manipulation element.

Then, how can Mele argue for premise 2? Consider his reasoning:

A defense of premise 2 might begin with the question how it can matter for the purposes of moral responsibility whether, in a deterministic world, a zygote with *Z*’s exact constitution was produced by a supremely intelligent agent with Diana’s effective intentions or instead in the way zygotes are normally produced. Imagine a deterministic world W^* that is a lot like the one at issue, W , but in which *Z* comes into being in Mary in the normal way and at the same time. It is conceivable that, in W^* , throughout his life, Mary’s child, Bernie, does exactly what Ernie does in W , down to the smallest detail.

Suppose that this is so, and suppose that Ernie and Bernie have matching

²⁶ The agnostics here are people who have spent time thinking about the question of free will and moral responsibility, but have not decided their answers. If the zygote argument can convince these people that compatibilism is false, then it is enough to put doubt into compatibilism.

“compatibilist-friendly agential structures,” Then, a proponent of ZAM might contend that, given the additional facts that, in both worlds, the featured agent has no say about what causes Z, no say about the rest of the universe at that time, and no say about what the laws of nature are, the cross-world difference in what caused Z does not support any cross-world difference in moral responsibility. In the same vein, a proponent of ZAM might be impressed by the thought that everything done by Ernie and Bernie is part of the unfolding of their initial conditions—that they are on equal footing on that score. (2019, 102-3)

The “no say” talk is the key. Mele claimed that both of Ernie and Bernie cannot affect their origins, their “initial conditions,” hence there is no significant difference justifying asymmetric judgments on moral responsibility of their actions. That is to say, from the perspective of Ernie and Bernie, they are equally powerless to the origins of the zygotes from which they developed. From this point of view, we should not view them as agents having different levels of control of themselves. Hence, they are equally responsible, or equally not responsible, for their behavior. Combined with premise 1, we get the conclusion that they are equally *not* responsible for their behavior.

4.4 Replying to the Zygote Argument

In this section, I will try to reply to the zygote argument, in order to defend my compatibilist position. First, I would like to spell out that I believe that premise 1 of ZAM is false. That is, Ernie *is* morally responsible for at least something he does.

According to Mele's teaching, the jury that we have to persuade is the agnostics, who have no prior commitment to compatibilism or incompatibilism. Hence, in what follows, I will avoid basing my argument on compatibilist or incompatibilist thoughts.

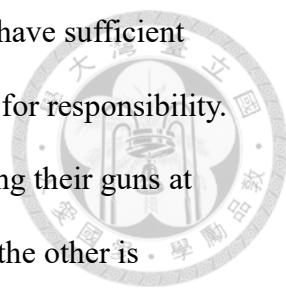
My argument starts from the thesis that responsibility is a matter of control. That is to say, whether one is responsible for something is decided by whether one has control over it.²⁷ For example, let us say that I run into a person in my car, and cause his injury. I am responsible for his injury only if the car is in my control. If I cannot control my car at all, then I am not responsible for his injury. I believe that *control* is the key element to understand what happens in Ernie's story. I would like to lay out my conclusion first: In my opinion, Mele neglects the control that Ernie has on his own actions.

4.4.1 Overlapping Control

At first sight, it might seem that Ernie is just Diana's tool, and that he completely lacks moral responsibility (and control) for doing anything. However, I believe that this is false. After a careful examination of the agency of Ernie, our intuition that Ernie is not morally responsible for anything he does may change.

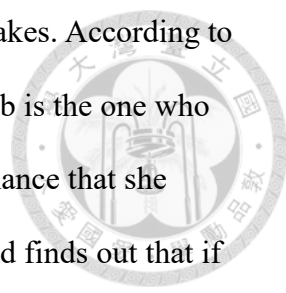
First of all, Diana has a strong control on the occurrence of event E. This is clear to see. With her omniscience and ability to create zygotes, she ensures that E will happen no matter what. However, the fact that Diana has a strong control on E does not necessarily rule out Ernie's control on E. This might seem weird at first sight. How can Diana and Ernie both have control on E? I call this kind of phenomenon

²⁷ I am only talking about direct responsibility here. For derivative responsibility, you also have to have control over some prior actions or events to be derivatively responsible.



“overlapping control,” in which more than one agent can be said to have sufficient control on the same event, sufficient in the sense of being sufficient for responsibility. Consider the following example: Two well-trained killers are pointing their guns at one person. One of the killers is pointing at the target’s head, while the other is pointing at his heart. Let us assume that a shot in this person’s head, and a shot in his heart are both a guarantee of his death. Now, both killers fire simultaneously. Let us also assume that the bullets fired penetrate the target’s organs simultaneously. After a short while, the target dies due to the injuries. In this case, it seems reasonable to say that both killers have equally sufficient control on the target’s death, and that both of them are responsible for his death. Their controls on the target’s death are equally sufficient, since either shot is sufficient for the death, which means neither killer’s responsibility is weakened by the presence of the other killer. We seem to have a case of overlapping control here: both killers have sufficient control on the target’s death, and neither control excludes the other’s control.

The above scenario might seem somewhat different than the situation of Diana and Ernie, since in the killers’ scenario, the two killers have *symmetric* controls, while Diana and Ernie seem to be in an asymmetric relation. Diana precedes Ernie on one causal chain, while the two killers initiate two different causal chains. Hence, let us consider another case of overlapping control, in which two agents are on the same causal chain, and both of them have sufficient control on one and the same event: Amy and her husband, Jacob, attend a birthday party of a friend. At the end of the party, there is one piece of cake left. Amy wants that piece of cake, but she prefers not to ask for it by herself. Amy has an extraordinary mind-reading power. By reading Jacob’s mind, she knows for sure that under this kind of circumstance, if Jacob has a chance to decide who should get the last piece of cake, he is definitely going to let



Amy have it. He is such a sweet guy, and he knows his wife loves cakes. According to the information that Amy gets from her mind-reading power, if Jacob is the one who gets to decide who can have the last piece of cake, there is a 99% chance that she would get the cake. Also, Amy has read every other guest's mind and finds out that if they are going to decide who should get the last piece of cake, it is not likely that they would let Amy have it. Hence, with the intention of getting the cake, she suggests letting Jacob decide who should take the last piece. As Amy has predicted, after a short thinking, Jacob really decides to let her have the last piece of cake. During Jacob's decision making, Amy did not give him any hints, or coerce or threaten him to make any decisions. He actually makes his own decision. In this scenario, it seems plausible to me to say that Amy and Jacob both have sufficient control on Amy's getting the cake. Amy controls the result by using her mind-reading power, and suggesting to let Jacob decide. She sets the initial conditions that once unfold, would lead to her desired result with a 99% chance. Although she is not the one who actually makes the decision, everything is in her plan. In a sense, Amy *uses* Jacob to achieve the desired result. Amy's control on the result is strong enough, if not perfect, for her to be held responsible for getting the cake. On the other hand, Jacob controls the result by actually making the decision, without being restricted by anyone else. He is just a free man who loves his wife. Hence, I think it is reasonable to say that Amy and Jacob are both responsible for Amy's getting the cake.

This scenario is more complicated than the killers' case, hence it requires more clarifications. Regarding responsibility, I believe that this scenario is similar to the killers' case. The two killers' controls on the victim's death are both sufficient for them to be held *fully responsible* for the victim's death. Being *fully responsible* for an event means that the agent is *as* responsible *as if* the other party is absent, or the other

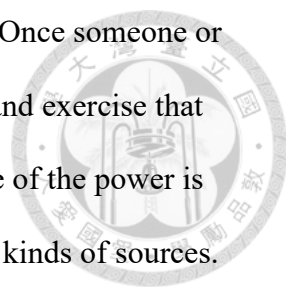
party has been substituted with something or someone that is *not* responsible for that event. For example, in the killers' case, both killers are *just as* responsible for the victim's death *as if* the other killer is absent, or *as if* we substitute a robot for the other killer. This means they are both *fully responsible* for the victim's death. In other words, they are *not* sharing the responsibility; instead, they are both *independently* responsible for the victim's death. Similarly, Amy and Jacob are both *fully responsible* for Amy's getting the cake: Amy is *as* responsible for her getting the cake *as if* we substitute anyone or anything that could play Jacob's role for Jacob, and Jacob is *as* responsible for Amy's getting the cake *as if* we substitute anyone or anything that could play Amy's role for Amy. In other words, they do not share the responsibility. Their responsibilities for Amy's getting the cake are *independent* to each other.

Let me elaborate on this point. Consider a hypothetical scenario, Robot: Instead of suggesting to let Jacob decide who should take the last piece of cake, Amy simply sends her robot to retrieve the cake. Amy is 99% sure that the whole process would work, and that she would get the cake. It seems to me that in the actual scenario and in Robot, Amy is equally responsible for her getting the cake. Since, no matter what means Amy uses to get the cake, as long as they are both 99% reliable, Amy's controls on the result would be equally firm. As long as Amy's controls on the result are equally firm, there would be no reason to assign different levels of responsibility to her in the two scenarios. Hence, Jacob does not affect the responsibility status of Amy regarding her getting the cake. Amy is *just as* responsible for her getting the cake *as if* we substitute a robot for Jacob, which could do Jacob's job but could not be held responsible for anything.

The situation of Jacob is analogous. Consider another hypothetical scenario, Draw: Instead of being appointed by Amy to make the decision, Jacob is randomly

drawn to make the decision. After being drawn, Jacob undergoes the same deliberation process as in the actual scenario, and makes the same decision. It seems reasonable to say that Jacob is equally responsible for Amy's getting the cake in the actual scenario and in Draw. Note that Amy's appointment does not have any bearing on Jacob's decision-making process. The fact that Jacob is appointed by Amy does not affect his decision at all. He would make the same decision no matter who tells him to make one. Hence, as long as Jacob has the power to make this decision by his own, no matter where this power comes from, his controls on the decision he makes, and the result he brings about, would be equally strong. Furthermore, as long as Jacob's controls on the result in the actual scenario and in Draw are equally strong, there is no reason to assign different levels of responsibility regarding the result to Jacob in the two scenarios.

Some might object that the fact that Jacob is appointed by Amy to make the decision *because* Amy has the intention to get the last piece of cake somehow *weakens* his control on the result, hence makes him *less* responsible for the result. However, I believe that this is false. First of all, *responsibility is a matter of control*. If Jacob is less responsible for the result when appointed by Amy than when randomly drawn, it must be because Jacob's control on the result is weaker when appointed by Amy than when randomly drawn. However, this is clearly not the case. When we are talking about Jacob's control on Amy's getting the cake, what is at issue is whether Jacob can *autonomously bring about* the result that Amy gets the last piece of cake, and this is true in both scenarios. Jacob would autonomously decide that Amy should get the cake, no matter appointed by who, or randomly chosen to make the decision. As long as the situation at the party remains the same, Jacob would have the same considerations, and would make the same decision. This shows that where Jacob's

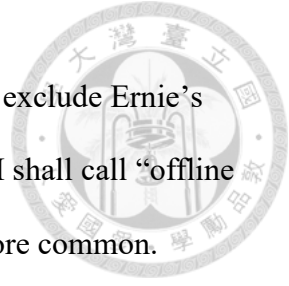


power comes from has nothing to do with his exercising the power. Once someone or something gives him the power to decide, he would be “activated” and exercise that power, not being affected by that someone or something. The source of the power is merely a “starter,” and Jacob’s exercising the power is neutral to all kinds of sources. Hence, the difference in the source of the power should not affect the strength of Jacob’s control on Amy’s getting the cake. In addition, if we focus on Jacob’s decision-making process and the result that it brings about, we would notice that the whole sequence in the actual scenario and that in Draw are identical. If the two sequences are identical, then it is reasonable to say that the levels of control they manifest are the same. This concludes my discussion of Jacob. I think that in the actual scenario, Jacob is *as* responsible for Amy’s getting the cake *as if* we substitute a draw for Amy. Hence, Jacob’s responsibility for Amy’s getting the cake is *independent* of Amy. I conclude that both Amy and Jacob are *fully responsible* for Amy’s getting the cake.

Now, we have another example of overlapping control. Both Amy and Jacob have sufficient controls on the result that Amy gets the last piece of cake. The controls they have are sufficient for both of them to be held *fully responsible* for Amy’s getting the cake. This is an *asymmetric* case of overlapping control.

Making clear that overlapping control is a real phenomenon is very helpful in discussing Ernie’s story, for I believe that Ernie’s story is actually a case of overlapping control. That is to say, both Diana and Ernie have sufficient control on event E. Hence, they are both morally responsible for event E. In the next section, I will try to argue for this idea.

4.4.2 The Adaptability of Control

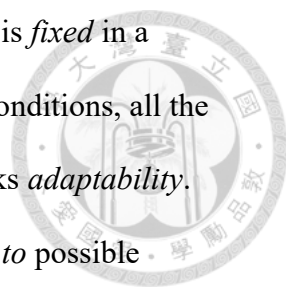


In what follows I will argue that Diana’s control on E does not exclude Ernie’s control on E. First, Diana’s control on E is of a special kind, which I shall call “offline control.” Contrary to offline control is “online control,” which is more common. When we are online-controlling something, our “acts of control” are *constant*. The acts of control are those acts we perform in order to make the desired result happen, or to ensure that no unwanted factors occur. For example, when controlling a robot, our acts of control might include pulling the lever to give orders to the robot. In the cases of online control, we “stay online” to constantly give orders to the object that we are controlling, or monitor and be prepared to give orders if necessary. We often online-control rather simple robots. We constantly give orders to control its movements: pulling the lever to make the robot go left or right. On the other hand, offline control is different. When we are offline-controlling something or some event, our acts of control are *fixed in a period of time*. The offline controller only sets the initial conditions that are needed to achieve the desired result, and then he “goes offline.” That is to say, the offline controller can somehow manage to ensure that the desired result will happen, merely by setting the initial conditions. Hence, after the initial conditions are set, the controller becomes completely idle. He lets the process unfold by itself, not intervening at all. One example of offline control is setting a time bomb. Suppose we set a time bomb in a stable environment where no one could find it, and make sure that there is no factor that might disable the bomb. After setting the bomb, we leave the scene and never return. In this scenario, we have control on the event that the bomb explodes, say, one hour later. Our control on the bomb’s explosion is one of offline style, since our acts of control (setting the bomb and checking the surroundings) are fixed in a period of time that ends one hour before the

desired result occurs. After that period of time, we go offline and do not perform any acts of control (not even monitoring). Diana's control on E is another example of offline control. With her omniscience, Diana is able to ensure that E will occur merely by creating Ernie's zygote. Once she has created the zygote, she completely goes offline and does nothing to Ernie. There is even no need to monitor and be prepared to intervene, since she already knows what will happen.

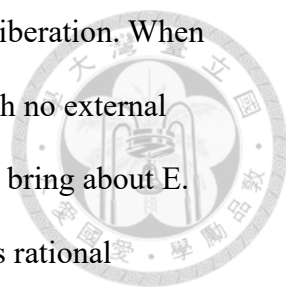
Let me summarize the common feature and the difference between online control and offline control. Both types of control exhibit the *sensitivity* of the result to the acts of control of the agent. The result is *sensitive* to the acts of control in the sense that, had the acts of control been some other ways, the result would be some other ways correspondingly. In the robot's example, if I pull the lever to the right rather than to the left, the robot would move to the right rather than to the left. Similarly, in the time bomb's example, if I set the bomb to explode two hours later rather than one hour later, the bomb would explode two hours later rather than one hour later. In both types of cases, if the result is *not sensitive* to the acts of control, then the agent's control would be gone: if the robot's movement is no longer sensitive to my pulling the lever, then I no longer control it; if the time bomb's explosion time is not sensitive to my setting, then it is not the case that I have control on it. However, the difference between the two types of control is the *duration* of the acts of control. The acts of control in an online control are constant. They last to the moment that the desired result occurs. On the other hand, the acts of control in an offline control are fixed in a period of time. They terminate before the desired result occurs. In a case of offline control, there is a period of time in which the controller does not perform any acts of control, but still has control on the desired result.

It is this special feature of Diana's control on E that makes room for Ernie's own

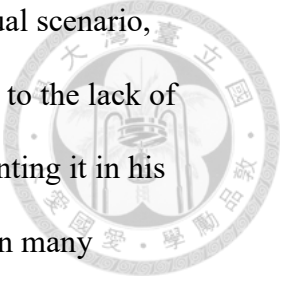


control. Unlike the constant nature of online control, offline control is *fixed* in a period of time. Since offline control consists only in setting initial conditions, all the works are *finished* after the conditions are set. In other words, it lacks *adaptability*. Saying that a controller is *adaptive* means that he is able to *respond to* possible changes. For example, I control my robot to walk across a street. Everything goes well. I simply make it walk straightly to cross the street. My control on it manifests *adaptability* if I am able to change my acts of control to respond to possible obstacles: if contrary to fact, there is a hole on the street, I would make it bypass the hole. Being able to use other strategies to achieve the same desired result in counterfactual scenarios means that in the actual scenario, my control on the result is *adaptive*. It is a feature of offline control that the controller would *lose* adaptability after the acts of control are finished. Since, if future conditions are not suitable for bringing about the desired result, the offline controller would not be able to change anything to revise them, for he has gone offline. As in the example of setting a time bomb, if contrary to fact, there is a malfunction in the bomb, I would not be able to fix it for I have left the scene. To be clear, adaptability is *not* necessary for an effective offline control. As long as the result is sensitive enough to the acts of control, it is a successful offline control. As in Diana's case, given Diana's full knowledge, she is able to make E happen or not happen by creating different kinds of zygote. Hence, adaptability is not necessary to make her plan work. However, adaptability is still of central importance in answering the zygote argument. In what follows, I will argue that it is this lacking of adaptability of Diana, plus some other settings of Ernie's story, that *prove Ernie's own control on E*.

According to Ernie's story, Diana creates Ernie's zygote in a way that it will develop into a person who satisfies all compatibilist conditions for free will. That is to

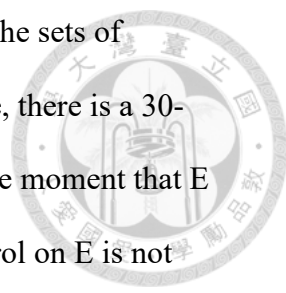


say, he is ideally self-controlled, and regularly exercises rational deliberation. When he decides to bring about E, he has evaluated many reasons, and with no external factors having coercive power on him, he concludes that it is best to bring about E. Now, judging from the *actual* scenario, it may still seem that Ernie's rational deliberation process is just a part of Diana's plan. However, I believe that the reason why we might have this kind of intuition is that we fail to pay enough attention to the *counterfactual* scenarios. Notice that Diana does not make Ernie a person who would decide to bring about E *no matter how the circumstance is*. Rather, she designs Ernie as a rational person, who would judge from his own reasons, as well as the environmental conditions that he is in, whether he should decide to bring about E or not. Looking into the counterfactual scenarios will make this point clearer. Suppose that E is the event in which Ernie kills his uncle. In the actual scenario, Ernie has many reasons that make him want to kill his uncle, and he is in a circumstance that is suitable for murdering his uncle. Let us say that Ernie's uncle always hits him when he was little, hence he hates his uncle badly. When Ernie is 30 years old, his uncle is drunk and tries to abuse him again. Now, Ernie is strong enough and has a knife near his hand. Considering all the above conditions, Ernie judges that it is the best moment to end his nightmare. Hence, he kills his uncle. However, in counterfactual scenarios, things would be different: If Ernie's uncle were way stronger than Ernie, then Ernie would reconsider whether attacking him is a wise choice; if Ernie did not have a knife near him, then he would only decide to defend himself; if when Ernie tried to kill his uncle, his uncle ran away, then he would chase his uncle. We can list many other counterfactuals. Although in the actual scenario, Ernie decides to bring about E, but there are many counterfactual scenarios where he would not decide to bring about E, or would decide to bring about E in different ways. The offline nature of Diana's



control partly explains why these counterfactuals are true. In the actual scenario, Diana's control is perfect, and E is ensured to happen. However, due to the lack of adaptability after her acts of control (creating Ernie's zygote and planting it in his mother) are done, Diana could not make E happen in the same way in many counterfactual scenarios. Diana is not able to "fix" possible changes, since she does not plan to intervene anymore. Thus, in many counterfactual scenarios, Ernie makes many other different decisions, bringing about many other different outcomes. Notice that not only does Diana lack adaptability, she also *endows Ernie adaptability*. She makes Ernie not a robot who would bring about E no matter what, but a person who is responsive to different conditions in the course of bringing about E, who can make different decisions according to different circumstances. I claim that these true counterfactuals exhibit Ernie's own control on bringing about E. They show that although Ernie's actions coincide with Diana's plan, he is not simply obeying Diana's orders, but actually making his own decisions. The fact that Ernie would decide to bring about many alternative events responding to many different counterfactual scenarios shows that in the actual scenario, he does have control on his decision and the event that he brings about.

I would like to talk more about the adaptability of *Diana's* control on E. To be clear, *before* Diana creates the zygote, her control on E is adaptive. If the future conditions were different, she would not create the zygote that she actually creates. She would create a different zygote that would give birth to a different person, who would bring about E in that different future. However, the adaptability of her control disappears after the moment she creates the zygote. It is the nature of offline control that it would lose its adaptability once the acts of control are done. After the acts of control are done, the controller can no longer respond to changes. Furthermore, the



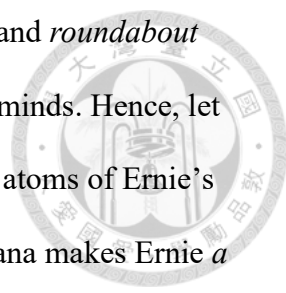
longer the period that the controller lacks adaptability is, the richer the sets of counterfactuals allowed for the controllee would be. In Diana's case, there is a 30-years gap between the moment that Diana creates the zygote, and the moment that E occurs. That is to say, in this very long period of time, Diana's control on E is not adaptive. Thus, very rich sets of counterfactuals are true for Ernie. I have listed a few, and here are some more examples: If Ernie's mother had protected him from his uncle's abuse, then Ernie would lack the motivation to kill his uncle; if the counsellor at school had known Ernie's situation and had taken care of his feelings, he would not become a person who would one day take murder into consideration; if Ernie had children when E was about to occur, he would not kill his uncle, since going to jail would leave his children without a father. These sets of true counterfactuals are so rich (thanks to the lacking of adaptability of Diana's control on E) that show that Ernie is not simply being controlled by Diana in the course of bring about E. Rather, he is a free agent who is making his own choices. When he decides to bring about E, all those true counterfactuals point to the fact that he is actually *in control of bringing about E*. He is not merely a puppet who would follow Diana's wish *no matter what*. There is a vast range of counterfactual scenarios in which he would choose to bring about many alternative events other than E. Hence, I conclude that the level of control that Ernie manifests is sufficient for him to be a suitable target of moral responsibility.

4.4.3 Explaining Away the Intuition

Now, I have explained why Ernie is an agent who has sufficient control on bringing about E. However, some of my readers might still be haunted by the intuition that Mele has mentioned. Mele claimed that upon hearing Ernie's story, the agnostics

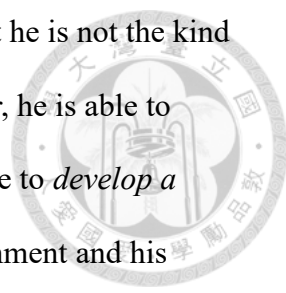
would have the intuition that Ernie is not morally responsible for anything he does, and that this intuition is caused by a fact F, “the fact that just by assembling the atoms of Ernie’s zygote as she does and implanting them in Mary when she does, Diana intentionally brings it about and intentionally ensures that Ernie will do $A_1 \dots A_n$, where that is a complete, ordered list of Ernie’s actions.” My job here is to explain away this non-responsibility intuition. I will try to explain why some of us might have this kind of false intuition.

First of all, F is indeed a fact. However, I believe that F is an oversimplified fact. F only tells us that “Diana intentionally brings it about and intentionally ensures” all of Ernie’s actions. It neglects *how* Diana brings about and ensures Ernie’s actions. F misses the *means* that Diana uses to make Ernie perform those actions. I guess for those audiences that have non-responsibility intuition, what they have in minds might be something like this: Diana wants some actions A, B, and C to be done by Ernie; hence, she uses Ernie as a tool; she gives orders to Ernie by creating his zygote, by which she *directly* causes Ernie to perform A, B, and C. If we construe Ernie’s story in this way, then of course we would get the non-responsibility intuition. However, this interpretation of the story is mistaken. Actually, the *means* that Diana uses to bring about and ensure Ernie’s actions is *making Ernie a person who can make decisions based on his own reasons*. This is the magic behind the story. The method that Diana employs is *not* simply giving orders to Ernie, programming him in a way that would make him do A at t_1 , do B at t_2 , and do C at t_3 . Rather, her method is to endow Ernie with the power of *rational deliberation*. She makes Ernie a person who would *deliberate* and *decide* to do A if the circumstance is such and such, and his reasons are such and such; *if* the circumstance and his reasons are some other ways, then he would *deliberate* and *decide* to do B or C. That is to say, the *means* that Diana



uses to bring about and ensure Ernie's actions is way more *indirect* and *roundabout* than what those audiences with non-responsibility intuition have in minds. Hence, let us specify F in a detailed and complete way: Just by assembling the atoms of Ernie's zygote as she does and implanting them in Mary when she does, Diana makes Ernie *a person with the power of rational deliberation, who can make decisions based on considerations of the circumstance and his own reasons*. By this method, Diana intentionally brings it about and intentionally ensures that Ernie will do $A_1 \dots A_n$, where that is a complete, ordered list of Ernie's actions. This is the complete form of F. It specifies the *means* by which Diana intentionally brings about and ensures Ernie's actions. My claim is, if the audiences understand F in its complete form, the number of them with non-responsibility intuition would drastically drop. This is because they get to realize that the *means* employed by Diana is one which turns Ernie into *a fully capable agent*, an agent who is capable of making his own decisions. Although it is still true that Diana intentionally brings about and ensures all of Ernie's actions, the non-responsibility intuition that this part of F provides would be countered by another intuition that the other part of F provides: a fully capable agent with the power of rational deliberation, who is able to make his own decisions based on his reasons and the considerations of the circumstance, is responsible for his actions. I am not claiming that the audiences would immediately have a clear intuition that Ernie is responsible for many of his actions, but I do believe that considering the complete form of F would balance out the original non-responsibility intuition for most of the audiences.

For some audiences, Ernie might seem like Diana's *tool*. This kind of thought deserves more discussions. In one sense, Ernie *is* Diana's tool, for Diana creates him in order to bring about some events. However, I believe what is being neglected here

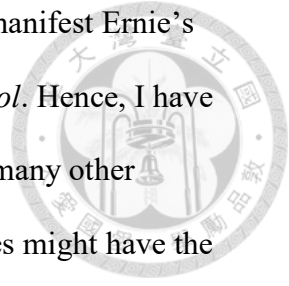


is that *Ernie is a very powerful tool*. He is powerful in the sense that he is not the kind of tool that simply receives orders and responds accordingly; rather, he is able to form his own decisions in an extremely sophisticated way. He is able to *develop a personality*, which comes from his earlier stages of growing environment and his earlier interactions with it. With this personality, he is able to make decisions based on it. Furthermore, he has many emotions, which also affect his decision making. As he grows older, he is able to learn to control his emotions and make sure they do not overtake his rationality during his decision-making processes. In the course of his life, he keeps collecting reasons based on which he will make later decisions. He is even able to develop desires, which sometimes make him unable to pick the most rational option. All the above show that if Ernie is a tool, then he is a very *autonomous* tool. Although his actions coincide with Diana's intentions eventually, his decision-making processes nevertheless exhibit a high level of *autonomy*. I believe that a person with this level of *autonomy* is indeed responsible for his actions, no matter he is a tool or not. In conclusion, if Ernie is a tool, then he is a very powerful and sophisticated tool that is an appropriate target of moral responsibility.

4.4.4 Summing Up

My reply to the zygote argument focuses on the premise 1 of ZAM: Ernie is not morally responsible for anything he does. I have argued that Ernie is indeed morally responsible for many of his actions. I have claimed that Ernie's story is a case of overlapping control, in which both Diana and Ernie have sufficient control on E, hence are both morally responsible for E. Ernie's control on E comes from a special nature of Diana's offline control on E: the lacking of adaptability. Diana's lacking of

adaptability allows abundant true counterfactuals for Ernie, which manifest Ernie's own control on bringing about E. *Responsibility is a matter of control*. Hence, I have concluded that Ernie is morally responsible for E, and analogously many other actions. Last but not least, I have also explained why some audiences might have the false non-responsibility intuition.



4.5 Conclusion

In this chapter, I have tried to defend compatibilism against the attacks from the manipulation arguments. I have replied to what I believe to be the most powerful form of the manipulation argument, Mele's zygote argument. If my reply to the zygote argument is successful, then I believe that it is safe to say that there is no serious threat to compatibilism, at least for the moment.

What can we learn from the zygote argument? If Ernie can maintain his control and free will under the creation of Diana, then it seems that free will is really tightly linked to the concept of a *person*. After all, even if Ernie is created by Diana in order to realize her plan, as long as Ernie is a *person* who has lived his own life, it seems that his agential behaviors would inevitably exhibit free will. This gives us more courage to face determinism. Even if our whole life is determined, or even if we are made by the intention of the Creator, as long as we can develop into a fully capable rational agent, we can be free.

Combining chapter 3 and 4, we can get a silhouette of source compatibilism. In the next chapter, more concrete contents will be filled in. I will provide a positive account of my source compatibilism, based on the lessons taught by Tim and Ernie.

Chapter 5



A New Source Compatibilism

5.1 Foreword

In this last chapter, I wish to propose a new type of source compatibilism based on my findings from the discussions of Brain Transformation and the zygote argument. It will be a source theory, since Brain Transformation has shown the falsity of PAP+; it will be a compatibilism, since the requirements that I am going to put forward are compatible with determinism.

5.2 Ernie and the Reasons-responsive Theory

In chapter 4, I have argued that Ernie is morally responsible for many of his actions because a look into the counterfactual scenarios shows that Ernie does have *control* in the actual scenario. Now, my job here is to formulate a theory that can articulate this *control* that Ernie manifests. First of all, this *control* crucial to free will and moral responsibility is one exhibited by the *actual causal sequence* of the agent.²⁸ We want to express the thought that in actually bringing about event E, Ernie uses his

²⁸ For another view focusing on the actual causal sequence of the agent, but with a much heavier emphasis on the causal properties of the sequence, see Carolina Sartorio (2016).

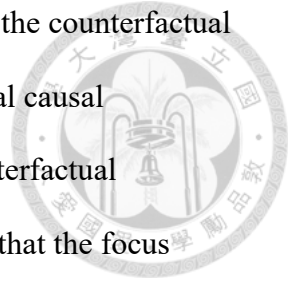
ability as a fully capable rational agent; having gone through a rational deliberation process, he finally arrives at his decision; *thus*, he manifests control. Secondly, I have claimed that this control on the actual causal sequence can be established by true counterfactuals describing Ernie's responses to different circumstances in the counterfactual scenarios. These two lines of thought, taken together, seem to be pointing at the familiar "reasons-responsive theory."

The most famous and sophisticated version of reasons-responsive theory is the one developed by Fischer and Ravizza in their (1998). Fischer and Ravizza claimed that for an agent to be morally responsible for his actions, the agent has to act on a mechanism that appropriately connects his actions to his reasons, and is able to recognize and react to possible reasons to do otherwise. (1998, 81-2) By "being able to recognize and react to reasons to do otherwise," Fischer and Ravizza meant that if sufficient reasons to do otherwise were present to the agent, the agent would have to be able to recognize some of them, and in some of the cases that the agent recognized the reasons to do otherwise, the agent would really act otherwise for those reasons.

It seems that Fischer and Ravizza's reasons-responsive theory can perfectly account for Ernie's moral responsibility for bringing about E. Firstly, on the actual causal sequence, Ernie's actions are connected to his reasons; he brings about E after a rational deliberation process in which he has considered his reasons. Secondly, in many counterfactual scenarios, Ernie satisfies the requirement of recognizing and reacting to reasons to do otherwise; in some counterfactual scenarios in which there were reasons not to kill his uncle (E), he would recognize them, and react to them by bringing about some other events. It seems that my job has been done: Fischer and Ravizza's theory has successfully articulated the *control* that Ernie manifests.

Unfortunately, this is not yet the end of story. Although I believe that the

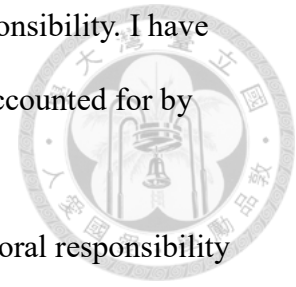
direction of Fischer and Ravizza's theory is correct, that a look into the counterfactual scenarios is required to bring out the *control* manifested by the actual causal sequence, however, I think they put too much emphasis on the counterfactual scenarios, thus making their requirements too demanding. I believe that the focus should be put on the *actual causal sequence*. The only reason that we have to look into the counterfactual scenarios is to bring out the *control* on the actual causal sequence. Hence, as long as we can bring out that *control*, the lighter the requirements on the counterfactual scenarios would be better. Moreover, Fischer and Ravizza's heavy focus on the counterfactual scenarios makes them rule out some free agents, free agents like Tim.



5.3 Lessons from Tim

Let us recap Brain Transformation quickly. Tim's brain is transformed; the functions of all the sequences of its neural activities have been changed into "making Tim fall into permanent sleep," except one special sequence of neural activities S, whose functions remain normal. Due to sheer luck, the sequence of neural activities that Tim performs in his life is exactly S, hence his brain functions normally throughout his whole life. My purpose is to make Tim an agent who has only one kind of alternative possibility, which is a kind of alternative possibility that is totally irrelevant to the actions that he actually performs. I have argued that Tim is morally responsible for many actions in his life, since the sequence of neural activities he performs has not been affected in any way by the brain transformation, nor is the performances of those neural activities forced upon him by it. Hence, we should not treat him differently from a normal moral agent. However, the alternative possibilities

that Tim has are too bizarre to be able to account for his moral responsibility. I have thus concluded from this case that moral responsibility cannot be accounted for by alternative possibilities in themselves.



By Brain Transformation, I want to show that the *ground* of moral responsibility is not the alternative possibilities the agent has; rather, what can account for one's moral responsibility is the *control* exhibited by one's *actual causal sequence*. As long as an agent's actual causal sequence exhibits control, the agent has free will and is morally responsible for his actions, even though he lacks the kind of alternative possibilities that are crucial for the leeway theorists.

What if we apply Fischer and Ravizza's reasons-responsive theory to Brain Transformation? The mechanism on which Tim acts clearly cannot recognize or react to reasons to do otherwise.²⁹ If the reasons to do otherwise were present to Tim, arguably he would perform some neural activities different from those of S, and fall into permanent sleep immediately. It is impossible for Tim to recognize those reasons, let alone reacting to them. Thus, according to Fischer and Ravizza's theory, Tim is not morally responsible for anything he does in his whole life.

However, I believe that this is false. There is a significant difference between Tim and a robot, or between Tim and a person who has no control on his behaviors. Fischer and Ravizza's theory fails to capture this difference. Merely because the mechanism on which Tim acts cannot respond to reasons, he has been put into the "no control" category. This is misleading. Tim does have control on his actions. His control on his actions can be shown by the following two facts. First, unlike a robot whose actions are caused by the controller's orders, Tim's actions are based on his

²⁹ In Fischer and Ravizza's terms, a mechanism is "the process that leads to the relevant upshot." (1998, 38)

own intentions and reasons. Second, were the circumstance to be different from what it actually is, Tim would not make the same decisions and perform the same actions like a robot. These two facts show that in actually performing his actions, Tim does have control on them.

I believe that the lesson taught by Brain Transformation is closely linked to the flaw in Fischer and Ravizza's theory: alternative possibilities are not the ground of moral responsibility. Although Fischer and Ravizza's aim is to develop a source theory, I believe that their requirements of reasons-receptivity and reasons-reactivity in the counterfactual scenarios have pushed their theory a bit too close to the leeway theory. Reasons-receptivity and reasons-reactivity are very *concrete* requirements for the agent in the counterfactual scenarios. They are *concrete* in the sense that they require the agent to really perform some specific actions (recognizing reasons and reacting to them) in the counterfactual scenarios. These requirements are so *concrete* that blur the distinction between them and the requirement of genuine alternative possibilities. It seems that the only difference between them is that one requires *hypothetical* alternative possibilities while the other requires *actual* alternative possibilities. This is why Brain Transformation can at the same time be a counterexample to PAP+ and to Fischer and Ravizza's reasons-responsive theory: the requirement of the latter is surprisingly close to that of the former. Thus, I believe that the true source theory should be less demanding on the counterfactual scenarios. Also, I am confident that a less demanding requirement on the counterfactual scenarios is still enough to establish the crucial *control* on the actual causal sequence.

5.4 A New Source Compatibilism

This is the right moment to put forward systematically my account of free will and moral responsibility. I believe that to have free will for one's actions, it is both *necessary* and *sufficient* for the agent to manifest *control* on his or her *actual causal sequence*; to be morally responsible for one's actions, this condition is also a *necessary* condition. The *control* on the actual causal sequence is to be manifested by the satisfying of the following condition, which I call the "*actual-control condition*." It has two clauses.

Actual-control condition

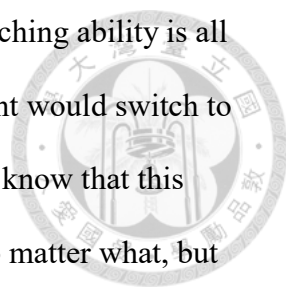
1. The actual-sequence clause:

The actions that the agent actually performs have to be appropriately connected to his reasons.

2. The counterfactual-sequences clause:

There exists a vast range of reasons R_1, \dots, R_n to do actions A_1, \dots, A_n , such that for any i , (i) it is reasonable for the agent to respond to R_i by A_i , and (ii) if R_i were present, the agent would not proceed with the causal sequence S that he actually proceeds with, but would *initiate* other causal sequence S_i which could lead to A_i .

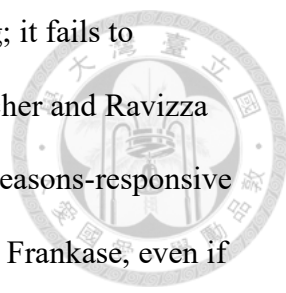
Readers can see that what is special in my actual-control condition are the requirements of "alternative causal sequence initiation" and "being reasonable for the agent." Let us start with the first feature. I believe that if the agent would *initiate* an alternative causal sequence were the reason to do otherwise to be present, then it is enough to show that he has *control* on proceeding with the actual causal sequence. I



call this kind of ability “sequence-switching.” I think sequence-switching ability is all that we require for an agent to have control on his actions. If an agent would switch to alternative causal sequences had reasons to do so occurred, then we know that this agent is not a robot who would stick to the same causal sequence no matter what, but a capable agent who has control on which causal sequence he is going to proceed with. If he actually proceeds with S, then we can know that he *chooses* to proceed with S, rather than being forced to proceed with S. It is at this point that I believe Fischer and Ravizza have made a mistake. The ability of sequence-switching does not have to specify “what would the agent do on the alternative causal sequence.” What we only need is that the agent would *initiate* an alternative causal sequence when necessary. Being able to initiate alternative causal sequence is enough to show that he has the ability of sequence-switching: If there were a need not to proceed with the original causal sequence, he would switch to other causal sequences.³⁰

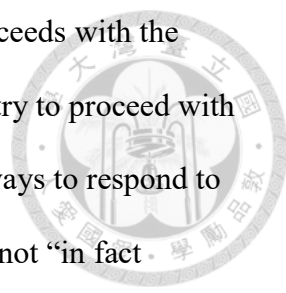
Let us move on to the second feature, that the actions that the counterfactual causal sequences would lead to should be reasonable ways to respond to reasons *for the agent*. Why do we need this requirement? It is because sequence-switching ability needs some restrictions. The way that the agent switches to alternative causal sequences must satisfy some patterns. It cannot be the case that the agent would randomly switch to any alternative causal sequences if there were a need not to proceed with the actual one. Thus, the alternative causal sequences that the agent would initiate if reasons to do so were present must have a special property, a property that can reinforce the agent’s control on the actual causal sequence. I have claimed

³⁰ My theory fits what Sartorio called “weakening-of-the-consequent strategy.” She criticized that some forms of this kind of strategy fail because their requirements are not robust enough to ground freedom. See Sartorio (2016, 118-9). I remedy this weakness of my theory by appealing to the second feature.



that Fischer and Ravizza's reasons-responsiveness is too demanding; it fails to account for Tim's control on his actions. In fact, the reason that Fischer and Ravizza have to appeal to a mechanism-based theory is that an agent-based reasons-responsive theory has problem explaining Frankases in general. Recall that in a Frankase, even if there were sufficient reasons to do otherwise, the agent would not react to those reasons by doing otherwise, since the counterfactual intervener would force him to do what he actually does. Thus, Fischer and Ravizza have to switch to a mechanism-based theory. (1998, 38) However, the concept "mechanism" itself has many difficulties.³¹ I think the real problem pointed out by the Frankases, and in specific by Brain Transformation, is what I have argued: Fischer and Ravizza's reasons-responsive theory is too demanding on the counterfactual scenarios. Hence, I believe that to accommodate the lessons from the Frankases including Brain Transformation, the right way is to be less demanding on the counterfactual scenarios, and keep our eyes on the actual causal sequence. However, some restrictions are still needed. Thus, let us go back to the "special property" that I have mentioned above. I believe that the alternative causal sequences that the agent would switch to should be able to lead to actions that would be reasonable ways to respond to reasons for the agent. This requirement helps manifest some aspects of the agent's rationality; it shows that if there were reasons to do otherwise, the ways that the agent would react would not be results of a random process. Rather, the agent would initiate a particular alternative causal sequence *because* it would be reasonable for him to respond to a particular reason by doing so. The satisfying of this requirement exhibits the agent's *control* on the actual causal sequence: In deciding to actually proceed with this particular causal sequence, the agent is using his rationality; it is because the actual causal sequence is

³¹ For detailed discussion, see Mckenna (2013).



a reasonable way to respond to actual reasons for him, hence he proceeds with the actual causal sequence; if there were some other reasons, he would try to proceed with (initiate) some other causal sequences which would be reasonable ways to respond to those reasons for him. But why do I say “reasonable for the agent,” not “in fact reasonable?” This is because we do not have to require that the agent would always initiate causal sequences which would *in fact* be reasonable ways to respond to reasons. We all sometimes have misunderstandings about the situation; what we think to be a reasonable way to respond to a reason might not in fact be one. However, this kind of misunderstanding does not hinder our exercising of rationality. Even if I have misjudged the situation, as long as the causal sequence that I choose to proceed with is a reasonable way to respond to a certain reason for me, I am still exercising my rationality, and that is all we need to exhibit *control* on our actions. In conclusion, the requirement that the causal sequences that the agent would initiate should be reasonable ways to respond to reasons *for the agent*, is an appropriate, and not too demanding, restriction of the agent’s sequence-switching ability.

My *actual-control condition* can explain why in Brain Transformation, Tim has control on his actions. First, Tim satisfies the actual-sequence clause. The actions that he actually performs are appropriately connected to his reasons. Second, Tim satisfies the counterfactual-sequences clause. Let us consider some counterfactual scenarios in which there were sufficient reasons to do otherwise. In these counterfactual scenarios, Tim would *initiate* other causal sequences different from the actual one. Although in doing so, he would fall into permanent sleep, this would not stop him from *initiating* those alternative causal sequences. This possible act of *initiating* alternative causal sequences proves his ability of *sequence-switching*. Furthermore, the alternative

causal sequences that he would initiate would be reasonable ways to respond to those reasons for him. Although those sequences would not in fact be reasonable ways to respond to reasons, since their functions would not be normal, they would still be reasonable ways to respond to reasons for Tim (he does not know he has undergone the surgery), and this proves his exercising of rationality in proceeding with the actual causal sequence. In sum, Tim satisfies the actual-control condition. Thus, his actual causal sequence manifests *control*, and that is why Tim is a free agent and a suitable target for moral responsibility. We can see that actual-control condition successfully captures the *control* that Tim has on his actions.³²

5.5 Closing Remarks

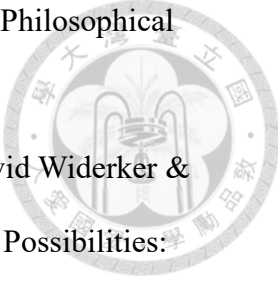
I have proposed a new source compatibilism, based on a key condition, the *actual-control condition*. My theory is motivated by a new blockage case that I have developed, Brain Transformation, and my treatment of Mele's zygote argument. During my construction of Brain Transformation, I have found out that the ground of free will and moral responsibility is not the availability of alternative possibilities. Thus, I have been focusing on the *control* that an agent manifests in the actual causal sequence. During my investigation into the zygote argument, I have reinforced my thought that the *control* that I have in mind is indeed compatible with causal determinism. A new form of source compatibilism is thus born.

³² Readers may try to apply actual-control condition to Ernie as well, and I believe would also get the right verdict.

References

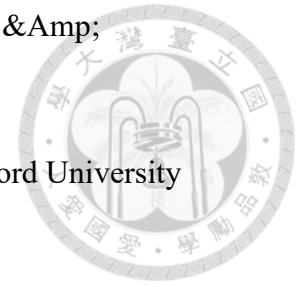


- Ayer, A. J. (1954). Freedom and necessity. In *Exploring Philosophy: An Introductory Anthology*. Palgrave Macmillan. pp. 271-284.
- Copp, David (2003). 'Ought' Implies 'Can', Blameworthiness, and the Principle of Alternate Possibilities. In David Widerker & Michael McKenna (eds.), *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities*. Ashgate. pp.
- Fischer, John Martin (1994). *The Metaphysics of Free Will: An Essay on Control*. Cambridge, Mass.: Wiley-Blackwell.
- Fischer, John Martin (1999). Recent work on moral responsibility. *Ethics* 110 (1):93-139.
- Fischer, John Martin (2006). *My way: essays on moral responsibility*. New York: Oxford University Press.
- Fischer, John Martin & Ravizza, Mark (1998). *Responsibility and Control: A Theory of Moral Responsibility*. New York: Cambridge University Press. Edited by Mark Ravizza.
- Frankfurt, Harry G. (1969). Alternate Possibilities and Moral Responsibility. *Journal of Philosophy* 66 (23):829-839.
- Frankfurt, Harry G. (1971). Freedom of the will and the concept of a person. *Journal of Philosophy* 68 (1):5-20.
- Ginet, Carl (1996). In defense of the principle of alternative possibilities: Why I don't find Frankfurt's argument convincing. *Philosophical Perspectives* 10:403-17.
- Haji, Ishtiyaque (1998). *Moral appraisability: puzzles, proposals, and perplexities*. New York: Oxford University Press.

- 
- Hunt, David P. (2000). Moral responsibility and unavoidable action. *Philosophical Studies* 97 (2):195-227.
- Hunt, David (2003). Freedom, foreknowledge, and Frankfurt. In David Widerker & Michael McKenna (eds.), *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities*. Ashgate. pp. 159--183.
- Hunt, David P. (2005). Moral responsibility and buffered alternatives. *Midwest Studies in Philosophy* 29 (1):126-145.
- Kane, Robert (1996). *The Significance of Free Will*. New York, US: Oxford University Press USA.
- Kane, Robert (2000). Responses to Bernard Berofsky, John Martin Fischer and Galen Strawson. *Philosophy and Phenomenological Research* 60 (1):157-167.
- Kane, Robert (2003). Responsibility, indeterminism and Frankfurt-style cases: A reply to Mele and Robb. In David Widerker & Michael McKenna (eds.), *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities*. Ashgate. pp. 91-105.
- Lewis, David K. (1976). The Paradoxes of Time Travel. *American Philosophical Quarterly* 13 (2):145-152.
- McKenna, Michael S. (1997). Alternative Possibilities and the Failure of the Counterexample Strategy. *Journal of Social Philosophy* 28 (3):71-85.
- McKenna, Michael S. (2001). Source incompatibilism, ultimacy, and the transfer of non-responsibility. *American Philosophical Quarterly* 38 (1):37-51.
- McKenna, Michael S. (2013). Reasons-Responsiveness, Agents, and Mechanisms. In Shoemaker (ed.), *Oxford Studies in Agency and Responsibility Volume 1*. New York: Oxford University Press. pp. 151-183.
- Mele, Alfred R. (2006). *Free Will and Luck*. New York, US: Oxford University Press.

- Mele, Alfred R. (2013). Manipulation, Moral Responsibility, and Bullet Biting. *The Journal of Ethics* 17 (3):167-184.
- Mele, Alfred R. (2019). *Manipulated Agents: A Window to Moral Responsibility*. New York, NY: Oup Usa.
- Mele, Alfred R. & Robb, David (1998). Rescuing Frankfurt-style cases. *Philosophical Review* 107 (1):97-112.
- Otsuka, Michael (1998). Incompatibilism and the avoidability of blame. *Ethics* 108 (4):685-701.
- Pereboom, Derk (2000). Alternative possibilities and causal histories. *Philosophical Perspectives* 14 (s14):119-138.
- Pereboom, Derk (2001). *Living Without Free Will*. New York: Cambridge University Press.
- Pereboom, Derk (2003). Source incompatibilism and alternative possibilities. In Michael S. McKenna & David Widerker (eds.), *Freedom, Responsibility, and Agency: Essays on the Importance of Alternative Possibilities*. Ashgate. pp. 184-199.
- Pereboom, Derk (2014). *Free Will, Agency, and Meaning in Life*. New York: Oxford University Press.
- Rocca, Michael Della (1998). Frankfurt, Fischer and flickers. *Noûs* 32 (1):99-105.
- Sartorio, Carolina (2016). *Causation and Free Will*. Oxford, United Kingdom: Oxford University Press UK.
- Stump, Eleonore (1988). Sanctification, hardening of the heart, and Frankfurt's concept of free will. *Journal of Philosophy* 85 (8):395-420.
- Stump, Eleonore (1996). Libertarian freedom and the principle of alternative possibilities. In Jeff Jordan & Daniel Howard-Snyder (eds.), *Faith, Freedom, and*

Rationality: Philosophy of Religion Today. Lanham: Rowman & Littlefield. pp. 73-88.



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

Widerker, David (1995). Libertarianism and Frankfurt's attack on the principle of alternative possibilities. *Philosophical Review* 104 (2):247-61.

Widerker, David (2000). Frankfurt's attack on the principle of alternative possibilities: A further look. *Philosophical Perspectives* 14 (s14):181-202.

Widerker, David (2003). Blameworthiness and Frankfurt's argument against the principle of alternative possibilities. In David Widerker & Michael McKenna (eds.), *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities*. Ashgate. pp. 53-73.

Widerker, David & McKenna, Michael (eds.) (2003). *Moral Responsibility and Alternative Possibilities: Essays on the Importance of Alternative Possibilities*. Ashgate.

Wolf, Susan R. (1990). *Freedom Within Reason*. New York: Oup Usa.

Wyma, Keith (1997). Moral Responsibility and Leeway for Action. *American Philosophical Quarterly* 34 (1):57 - 70.