ethic@
http://dx.doi.org/10.5007/1677-2954.2019v18n3p285

TRANSFERRING NON-RESPONSIBILITY

PEDRO MERLUSSI[1]
(UNICAMP/Brasil)

GABRIEL DE ANDRADE MARUCHI[2]
(UNICAMP/Brasil)

ABSTRACT

The Direct Argument argues for the claim that determinism and moral responsibility are incompatible. The most controversial assumption of the argument is the thought that "not being responsible for" transfers across conditionals: if p is true and no one is (even partially) morally responsible for the fact that p is true, and $p \supset q$ is true and no one is (even partially) morally responsible for the fact that $p \supset q$ is true, then q is true and no one is (even partially) morally responsible for the fact that q is true. Here we argue that the principle is true if one accepts a truth-maker account of the relationship between non-responsibility and propositions. While non-responsibility transfers across conditionals, one upshot of the truth-maker account is that it allows one to be responsible for necessary truths.
Keywords: Moral responsibility; Incompatibility; Ethics

Introduction

One of the main problems in current discussions of moral responsibility is answering the question "Does determinism rule out moral responsibility?"'. Determinism is the thesis that the past and the laws of nature determine a unique future (Earman (1986), van Inwagen, (1983)). Incompatibilism is the conditional thesis that if determinism is true, there is no free will in the sense that is relevant for moral responsibility. There are two basic ways for arguing for incompatibilism. One is indirect, such as the Consequence Argument, which argues for the claim that determinism rules out alternative possibilities. This is an indirect strategy because even if the argument were sound, one would also need to argue for the claim that

(PAP) If no one is ever able to do otherwise, then no one is morally responsible.

(PAP) stands for the principle of alternate possibilities. The basic thought is that morally responsible agents can be excused for acting wrongly if they are unable to do otherwise.

However, the obvious problem with the indirect strategy is that (PAP) is hugely controversial. Frankfurt (1969) challenged the principle in a way that his argument is compared to Gettier's counterexamples to knowledge as justified true belief (Vihvelin, 2013: 93), though there is the difference is that it is still controversial whether he provided a successful refutation of (PAP).

Fortunately for incompatibilists there is also the *direct* way. Rather than getting into too many controversial issues, the incompatibilist can give an argument without presupposing any necessary condition for moral responsibility. It need not use (PAP) as a premise, for example. So the direct way seems to have significant dialectical advantages over the indirect one. How such an argument could be constructed without resting on any claims about what moral responsibility requires? One simple strategy is to model an argument after the original Consequence Argument. This is the Direct Argument (van Inwagen 1983).

The Direct Argument makes use of a modal operator, "*M*". "*Mp*". is true just in case "*p* is true and no one is, or ever was, even partially morally responsible for the fact that *p*". Just like its counterpart, the Consequence Argument, the Direct Argument makes use of two deductive rules to argue for the claim that determinism and moral responsibility are incompatible. In the original formulation, van Inwagen proposed the following deductive rules:

A. $\Box\varphi \vdash M\varphi$

B. $M(\varphi \supset \psi), M\varphi \vdash M\psi\$$

Let *L* stand for the conjunction of the laws of nature and P_0 for a true proposition about the total state of the world in the past[3]. Finally, let *P* stand for any true proposition about human action. Here is the Direct Argument:

1.	$\Box((L \wedge P_0) \supset P)$	determinism
2.	$\Box(L \supset (P_0 \supset P))$	modal logic, 1
3.	$M(L \supset (P_0 \supset P))$	A, 2
4.	ML	premise
5.	MP_0	premise
6.	$M(P_0 \supset P)$	B, 3, 4
7.	MP	B, 5, 6

Perhaps the most controversial premise of the argument is B, the thought that "not being responsible for" transfers across conditionals, which is highly controversial (Ravizza, 1994). Here, however, we aim to show that B isn't that implausible. We argue for a novel way to understand the transference of non-responsibility by means of truth-makers and actions, which allows us to offer a new principle immune to Ravizza's objection to B. What is more, we show that if our principle holds, so does B. Though this essay may be seen as a defence of the transference of non-responsibility, it is not our aim to defend the Direct Argument. Quite the contrary, one upshot of our main proposal is that, while B goes through, A fails.

Here is the strategy. Section 1 presents Ravizza's alleged counterexample to B. Then in section 2 we argue for a new principle connecting non-responsibility and truth-makers, one that is immune to Ravizza's challenge to B. Having motivated our proposal, we argue that, if one accepts our claim, one ought to reject A. Finally, we discuss some possible objections to our account in section 3. The upshot of our essay is that, curiously enough, the reasons presented for B will also undermine A and, consequently, the Direct Argument.

1. Frankfurt-style scenarios as counterexamples to B

Frankfurt-style scenarios may be used to motivate counterexamples to B. Ravizza for instance gives the following scenario:

> **Erosion:** "Imagine that Betty plants her explosives in the crevices of a glacier and detonates the charge at T1, causing an avalanche that crushes the enemy fortress at T3. Unbeknownst to Betty and her commanding officers, however, the glacier is gradually melting, shifting, and eroding. Had Betty not placed the dynamite in the crevices, some ice and rocks would have broken free and crushed the enemy base camp at T3" (Ravizza 1994: 72–73)

To see how the alleged counterexample goes, let A stand for the proposition that there is an avalanche that crushes the enemy base at t_3 and E for the proposition that the glacier is eroding.

1. ME
2. $M(E \supset A)$
3. MA

What motivates 1 is the assumption that the fact the glacier is eroding is something inevitable. That is, no matter what anyone does, the natural forces would ensure that the glacier erodes. With respect to 2, Ravizza suggests that because the event *E* describes is a sufficient cause of the event that *A* describes, no one is even partially morally responsible for the material conditional:

> The idea here is that if there are conditions present (for which an agent is not responsible) and these conditions are sufficient to ensure a given outcome, then it seems as if the agent should not be held responsible for that outcome. (Ravizza 1994: 68)}

However, 3 seems false, for "Betty acts freely, and she is responsible for the consequences of her action" (1994: 72).

Ravizza's argument works in a similar way to Frankfurt-style scenarios because it is based on the distinction between the inevitability factors and the explanatory factors[4]. In *Erosion*, the fact that there is an avalanche that crushes the enemy base at t_3 is *inevitable*. The inevitability factors are the *pre-emptive potential cause*s of the avalanche[5]. What this means is that, even if Betty were not to plant her explosives, *A* would still be true, since the natural forces are playing the role of a preempted potential cause. But the inevitability factors are not playing a role in bringing about the avalanche. Nor are they playing a role in Betty's action; They do not explain why Betty planted her explosives. The scenario is such that the inevitability factors and the explanatory factors come apart, since the former does not explain why Betty acted the way she did[6].

All seems well and good, but is the counterexample really successful? We think not, as we shall argue in the next section.

2. The Partial meaning of "not being responsible for a proposition"

Suppose that *A*'s truth were to be grounded *only* by the natural forces, not by Betty's act. If *A* were true *because of* the natural forces, Betty would not be morally responsible. She would not be morally responsible because *A*'s truth would not be up to Betty, but to forces beyond her control. This is central to Ravizza's point. What is doing the explanatory work with respect to *A* is Betty's acting freely, rather than the inevitability factors. To see why, imagine that one were to interpret "not being responsible for" as follows (Huemer, 2000; van Inwagen, 2008):

(CS) *MP* is true iff *P* and, no matter what anyone does, *P* would still be true.

If (CS) captures the relevant sense of "not being responsible for", Betty will not be responsible for *A*, since *A* is resilient under counterfactual assumptions (that is, *A* would be true in that scenario "no matter what", since the natural forces are a preempted potential cause of the avalanche). Thus, if *Erosion* is a counterexample to rule B, (CS) cannot capture the relevant sense of "not being responsible for".

In what sense, then, is Betty at least partially morally responsible for the fact that *A* is true? Counterfactuals will not help us here because the inevitability factors and the explanatory factors come apart in the scenario. What this suggests, we think, is that the notion of being at least partially morally responsible for a proposition goes hand in hand with the agent's act explaining the truth of the proposition in question. We think this is the most plausible way to motivate Ravizza's claim that Betty is morally responsible for the fact that *A* is true.

So, the idea of our approach is to explain the intuition that Betty is morally responsible for *A* in terms of an *explanatory* notion. We offer just a *partial* analysis of "not being responsible for", and it is merely partial because it is widely assumed that moral responsibility has two conditions. One is metaphysical: You cannot be held accountable for what you did if your action was not *up to you* or if you did not act *freely*. The other condition is epistemic: You cannot be held accountable for what you did if you did not know what you were doing or if you were not aware of the moral significance of what you were doing. Our analysis aims to capture only the metaphysical condition (though we shall say something about the epistemic one later on).

(ES) *s* is (at least) partially morally responsible for the fact that *P* iff either the truth-maker of *P* is *s*'s free action or caused by her free action.

Notice that (ES), contrary to (CS), can do the job of capturing the thought that Betty is morally responsible for the fact that *A*. True, *A* would still be true even if Betty were not to detonate her explosives. However, the inevitability of *A* does not undermine Betty's moral responsibility because it is not part of the chain of events leading to the avalanche. Her detonating the charge at t_1 is. Betty performs an action that leads to the avalanche that crushes the enemy base at t_3.

2.1 Reply to Ravizza

Suppose for a moment that this partial analysis is correct. Suppose that the relationship between partial moral responsibility and propositions is cashed out (at least more or less) in terms of (ES). There is a simple consequence of this view. While (ES) can explain why Betty is partially responsible for the avalanche, it also explains why she is partially responsible for the fact that E materially implies A. If so, Ravizza's argument against B will fail!

We think Betty is (at least) partially responsible for the fact that E materially implies A given our partial analysis. To see why, we must first notice that $E \supset A$ is equivalent to $\sim E \vee A$. So, if Betty is morally responsible for the fact that $\sim E$ or A, she is morally responsible for the fact that E materially implies A. Now, (ES) will tell us that Betty is partially responsible for the fact that $\sim E$ or A iff either the truth-maker of $\sim E \vee A$ is Betty's act or caused by her act. If we want to know whether Betty is responsible for the disjunction, we first need to know what is the truth-maker of $\sim E \vee A$. Since this disjunction has only one true disjunct, namely A, then whatever makes A true also makes the disjunction true (Restall, 1996)). Taking this into account, following (ES), we have that Betty is (at least) partially morally responsible for the disjunction, since her detonating the charge at t_1 is a cause of the avalanche crushing the enemy base at t_3 (which is the truth-maker of A) (see also Hermes (2013)). Therefore, 2 is false and *Erosion* fails as a counterexample to B.

Of course, a similar strategy may be employed to defend B (leaving epistemic worries concerning moral responsibility aside). Notice that Betty is responsible for $E \supset A$ because (i) she is responsible for the fact that A is true and (ii) $E \supset A$ and A have the same truth-makers. To put it in a more schematic form, it goes as follows: (TM) if φ is true and s is partially responsible for φ, φ and ψ are true and have the same truth-makers, s is partially responsible for ψ.

If one accepts (TM), then one should also accept B. Notice that if B fails, there must be a situation in which some agent is partially responsible for Q, even though no one is or ever was partially responsible for P and $P \supset Q$, that is, MP and $M(P \supset Q)$, but $\sim MQ$. Since M is factive, both P and $P \supset Q$ are true, which means that Q is also true. But since P is true and $P \supset Q$ is equivalent to $\sim P \vee Q$, whatever makes Q true also makes $\sim P \vee Q$ true. If some agent is responsible for Q, then the truth-maker of Q is either the agent's act or caused by her act. Hence, if the agent is responsible for Q and Q has the same truth-maker of $\sim P \vee Q$, then the

agent is responsible for $\sim P \vee Q$, which contradicts the assumption that $M(P \supset Q)$. Hence, B is valid.

2.2 B is valid, but A isn't

While the partial analysis gives good reasons for accepting B, it also undermines the plausibility of A. What A tells us is that no one is even partially morally responsible for necessary truths. The fact that A fails may strike one as counter-intuitive, but it is just a simple consequence of our explanatory analysis of "not being responsible for".

The main reason why A fails stems from the fact that, when a disjunction has only one true disjunct, then whatever is responsible for the disjunct is also responsible for the disjunction (Hermes 2013). This principle clearly follows from (ES). Now, to illustrate why A fails, consider the following example given by Hermes (2013). Suppose that Mary murdered Matt. Consider the proposition expressed by "Either Mary murdered Matt, or she did not". What makes this proposition true is the fact that Mary murdered Matt. So, since Mary is responsible for the only true disjunct, she is also responsible for the disjunction. Of course, "Either Mary murdered Matt, or she did not" is necessarily true, and Mary is responsible for it.

3. Objections

We have defended two controversial theses so far, namely (i) the transference of non-responsibility across material conditionals, and (ii) the failure of principle A. We argued for (i) and (ii) based on the partial analysis encapsulated by (ES), which is an explanatory account of non-responsibility. In this section, we respond to two objections.

3.1 Borderline

Robinson (2016) presents the following scenario:

> **Borderline:** Suppose that, in the middle of the night, I kidnap you from your bed, put you on a plane, take you up to 30,000 ft., and throw you out the side while directly over the Arizona-Nevada border. The following disjunction is now true: Either you are going to fall to your death in Arizona, or you are going to fall to your death in

> Nevada. Which of these disjuncts is (or ends up being) true depends on the movements and positioning of your body as you fall and is entirely up to you. Suppose that somewhere around 20,000 ft. A glimpse of the Las Vegas lights reminds you of a promise you made to your dying mother never to set foot on Arizona soil again, and you position your body to head toward Sin City. (Robinson 2016: 1337)

How could one use this scenario to argue against (ES)? Let *R* stands for the proposition that you are going to fall to your death in Arizona and *N* for the proposition that you are going to fall to your death in Nevada. The truth-maker of *N* is caused by your action. Therefore, according to (ES), you are morally responsible for *N*. Now, consider the true disjunction $R \vee N$. Given that *R* is false, this dinjunction is made true by *N*. So, the truth-maker of *N* is also the truth-maker of $R \vee N$. Following (ES), you are morally responsible for $R \vee N$. This conclusion, however, is counter-intuitive. We would not regard you as morally responsible for the fact that ''Either you are going to fall to your death in Arizona, or you are going to fall to your death in Nevada''.

Reply: This is a very tricky case, and tricky cases usually require distinctions. In order to reply to Robinson, we need to draw a distinction between truth-functional disjunctions and non-truth-functional disjunctions (de Sa, (2009), Taasiewicz (2013)). A disjunction is truth-functional iff it is true only when one of its disjuncts is true. This is the disjunction we usually have in mind when using the symbol $\vee$. However, de Sa points (2009) out that not every disjunction is truth-functional. Disjunctions concerning vagueness and open future are not truth-functional. For instance, suppose that Nina tosses a coin, and suppose that the side that will land is defined by an indeterministic process. While the coin is in the air, it is true that "the coin will land tails or it will land heads". Nonetheless, since the coin has not landed yet, this disjunction is not truth-functional, given that it was not made true by any of its disjuncts. The truth-maker of a truth-functional disjunction is the truth-maker of its true disjunct. The truth-maker of a non-truth-functional disjunction is distinct from the truth-maker of its disjuncts.

If we accept this, then Robinson's example will not work. The disjunction he presents is clearly a case where the future is open, as it were. The truth-maker of this disjunction - Either you are going to fall to your death in Arizona or in Nevada - is distinct from the truth-maker of its disjuncts. Since it is not any of your actions that make the disjunction true, you are not responsible for it.

Moreover, (ES) can explain why Robinson is responsible for the fact that either you fall to your death in Arizona or in Nevada. For his kidnapping you is a cause of the truth-maker of this disjunction.

3.2 Bomb-making

Another objection to (ES) can be made when we take the epistemic condition for moral responsibility into account. You walk past your office and notice a squeaky button. Unbeknown to you, the button is connected to a bomb that explodes the nearby neighbourhood. Suppose you press the button, thus killing lots of innocent people. Clearly, you are not responsible for this, and (ES) does not take it into account, since it leaves the epistemic condition out.

Is there a way to accommodate the epistemic condition? Here is one suggestion:

(ES*) *s* is (at least) partially morally responsible for the fact that *P* iff (i) either the truth-maker of *P* is *s*'s free action or caused by her action and (ii) *s* knows that the truth-maker of *P* obtained because of *s*'s free action

This is a way to solve the problem since you do not know that your action is the truthmaker of the proposition that the bomb explodes. While this would solve this specific problem, another one arises. Suppose we have two propositions, *P* and *Q*, and suppose that they have the same truthmaker *T*. Now, imagine that your action were to cause *T* while satisfying the epistemic condition presented in (ES*). If this is the case, according to (ES*), *s* is morally responsible for *P* and *Q*, which generates the following case against (ES*):

> **Bomb-making:** Anne is a great electronic engineer. She is hired to fix a problem in a circuit of a governmental device. She is informed that the device is a medical equipment used in complex surgeries. However, that is a lie; the device is a bomb that will be used against innocent people. Anne does not know about medicine or bomb-making enough to discover that she was deceived. She successfully fixes the circuit and, therefore, the bomb. [7]

Intuitively, it would be acceptable if someone praised Anne for fixing the circuit. After all, she was responsible for doing it. Given that, it is correct to say that she is morally responsible for the fact that she fixed the circuit. On the other hand, it would be inappropriate to hold Anne morally responsible for making a bomb. She did not know it was a bomb! However, one might argue, her fixing the circuit and making the bomb have the same truthmaker. According to (ES), she would be (at least) partially morally responsible for making

the bomb. Therefore, given that she is not even partially morally responsible for making the bomb, (ES*) is false.

Reply: One crucial feature for the above example to work is the assumption that Anne does *one thing* under different descriptions. That is, the example will work only if one assumes the identity thesis concerning the individuation of acts. Davidson writes:

> I flip the switch, turn on the light, and illuminate the room. Unbeknownst to me I also alert a prowler to the fact that I am home. Here I do not do four things, but only one, of which four descriptions have been given (Davidson, 1963: 696)

Anne (1) fixes the problem in the circuit *and* (2) she fixes the bomb. Assuming the identity thesis, she is performing just one act under different descriptions. If we accept this and (ES*), Anne will be responsible for making the bomb, which is quite implausible.

What is interesting about this point is that it is often assumed that the discussion concerning the individuation of action has "little or no bearing on other issues" (Schlosser 2019: section 3.4). However, it does play an important role here. The above objection can be avoided if one adopts a fine-grained view on the individuation of actions. According to fine-grained views, how many actions you perform depends on how many act-properties are instantiated. So, for example, the property of (i) one's fixing the circuit is different from the property of (ii) one's fixing the bomb. For this reason, when Anne fixes and the circuit and when Anne fixes the bomb, she exemplifies two different act-properties. These two particular acts are then two different *act-tokens*, where an act-token is the exemplifying of a property by an agent (Goldman, 1970: 10). This allows us to provide the following amendment in our account:

(Amended ES) *s* is (at least) partially morally responsible for the fact that *P* iff either the truth-maker of *P* is *s*'s free act-token or caused by her free act-token and (ii) *s* knows that the truth-maker of *P* obtained because of *s*'s free act-token.

Anne is responsible for fixing the circuit because her act-token (her exemplifying the act-property of fixing the circuit) is the truth-maker of the proposition that she fixes the circuit, and this is something she knows. On the other hand, she is not responsible for fixing the bomb simply because she does not know that she is performing the act-token of fixing the bomb.

4. Final remarks

We have argued for a truth-maker account of non-responsibility. While the account has the upshot that some agents may be morally responsible for necessary truths (which is in line with some cases in the literature), it also shows that principle B is valid, so that non-responsibility transfers across conditionals. Two interesting points are worth highlighting. The first one is that non-responsibility transfers across material conditionals if one accepts the truth-maker account of non-responsibility, and the Direct Argument fails because of A is false, which is curious since A is typically considered less controversial than B. Second, the truth-maker account of non-responsibility we offered goes more in line with a fine-grained account of the individuation of actions, which seems to have some advantage over the identity thesis.

Acknowledgements: Funding was provided by Fundação de Amparo à Pesquisa do Estado de São Paulo - FAPESP - (Grants No. 2017/20532-8 and 2019/12207-5). Many thanks to Francesco Maria Ferrari, Rodrigo Cid, Gabriel Andersen Eugênio, Ariston Sócrates Guimarães Soares Cruz, Renato Semaniuc Valvassori and two anonymous referees for comments and suggestions.

Notes:

[1]University of Campinas (UNICAMP), Center for Logic, Epistemology and The History of Science, Campinas, S.P., Brasil. E-mail: p.merlussi@gmail.com

[2] University of Campinas (UNICAMP), Campinas, S.P., Brasil. E-mail: gabrielmaruchi@outlook.com

[3] 'Before any agents existed. See Campbell (2007).

[4] We are not going to reproduce Frankfurt's case against (PAP) since it can be found everywhere in the literature. See Sartorio (2017) for a nice survey.

[5] Here is an example to illustrate the case of preemption. Suppose that two people - X and Y - decide to break the window of your house, agreeing that one or other will throw a stone in the window. Acting side-by-side, X and Y go together to throw a stone and break the window at your place. Person X however throws the stone and succeeds in breaking the window, and Y gives up from throwing the stone as he sees X doing it. Person X is responsible for causing the window breaking event, while Y's actions are a preempted potential cause.

[6] We have presented *Erosion* as a case of causal pre-emption. But even if the example is thought of as a case of overdetermination, we agree with Ravizza that Betty is still morally responsible.

[7] *Bomb-making* is similar to the Zombie Case formulated by Turner (2014). However, *Bomb-making* has two advantages over the *Zombie Case*. First, differently from the Zombie Case, *Bomb-making* is a counterexample to (ES) and (ES*). Second, the arguments Turner presented against the *Zombie Case* do not work against *Bomb-making*.

References

Campbell, J. K. (2007). Free will and the necessity of the past. *Analysis* 67 (2), 105–111.

Davidson, D. (1963).Actions, reasons, and causes. *Journal of Philosophy* 60 (23), 685–700.

de Sa, D. L. (2009).Disjunctions, conjunctions, and their truthmakers. *Mind* 118 (470), 417–425.

Earman, J. (1986). *A Primer on Determinism.* D. Reidel.

Frankfurt, H. G. (1969). Alternate possibilities and moral responsibility. *Journal of Philosophy* 66 (23), 829–839.

Goldman, A. I. (1970). *A Theory of Human Action*. Princeton University Press.

Hermes, C. (2013). A Counterexample to A. *Philosophia* (United States) 42 (2), 387–389.

Huemer, M. (2000). Van inwagen's consequence argument. *Philosophical Review* 109 (4), 525–544.

Ravizza, M. (1994). Semi-compatibilism and the transfer of non-responsibility. *Philosophical Studies* 75 (1-2), 61–93.

Restall, G. (1996). Truthmakers, entailment and necessity. *Australasian Journal of Philosophy* 74 (2), 331–340.

Robinson, M. (2016). Truthmakers, Moral Responsibility, and an Alleged Counterexample to Rule A. *Erkenntnis* 81 (6), 1333–1339.

Sartorio, C. (2017). Frankfurt-style examples. In G. Meghan, T. Kevin, and L. Neil (Eds.), *Routledge Companion to Free Will*., pp. 179–190. Routledge.

Schlosser, M. (2019). Agency. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2019 ed.). Metaphysics Research Lab, Stanford University.

Talasiewicz, M., J. Odrow-Sypniewska, W. Wciórka, and P. Wilkin (2013). Do we need a new theory of truthmaking? some comments on disjunction the- sis, conjunction thesis, entailment principle and explanation. *Philosophical Studies* 165 (2), 591–604.

Turner, P. R. (2014). Truth and moral responsibility. In F. B. M. Dell'Utri and S. Caputo (Eds.), *New Advances in Causation, Agency, and Moral Responsibility*. Cambridge Scholars Press.

van Inwagen, P. (1983). *An Essay on Free Wil*l. Oxford University Press.

van Inwagen, P. (2008). The consequence argument. In P. V. Inwagen and D. W. Zimmerman (Eds.), *Metaphysics: The Big Questions*. Blackwell.

Vihvelin, K. (2013). *Causes, Laws, and Free Will: Why Determinism Doesn't Matter*. Oup Usa

Received/Recebido: 12/11/2019
Approved/Aprovado: 29/01/2020