

Chapter 7

The Direct Argument for Incompatibilism

David Widerker and Ira M. Schnall

1. Introduction

Traditionally, incompatibilists have employed the following argument to show that determinism is incompatible with moral responsibility (henceforth the “Traditional Argument”):

- (1) If determinism is true, then no agent could have avoided acting as he did. (The incompatibility of avoidability with determinism, IVD)
- (2) An agent is morally responsible for acting as he did only if he could have avoided acting as he did. (Principle of Alternative Possibilities, PAP)
- (3) Therefore, if determinism is true, then no agent is morally responsible for his acts.

In the early nineteen-eighties, Peter van Inwagen ([1983](#)) proposed another argument for the same conclusion, one that argues “directly” for it from some general and *allegedly uncontroversial* assumptions about moral responsibility and determinism (pp. 182–188).

Let “PAST” stand for a true proposition describing the state of the world at some time before the existence of the human race, “LAWS” stand for the conjunction of the laws of nature, and “PRESENT” stand for any true proposition about the state of the world now. Assume now the validity of the following two rules of inference:

$$(A) \Box p \vdash \text{NR}(p)$$

$$(B) \text{NR}(p), \text{NR}(p \supset q) \vdash \text{NR}(q),$$

where “ \Box ” represents broadly logical necessity, “ \vdash ” stands for “therefore,” “NR(p)” abbreviates “p and no one is (now), or ever has been even partly morally responsible for the fact that p,”¹ and “ \supset ”

stands for material implication. Let us call Rule (B) the “Transfer of Non-Responsibility Rule” or “Transfer NR” for short.² Van Inwagen’s argument may then be formulated as follows:

- (1) Assume that the thesis of determinism is true.
- (2) $\square [(PAST \ \& \ LAWS) \ \supset \ PRESENT]$ From 1
- (3) $\square [PAST \ \supset \ (LAWS \ \supset \ PRESENT)]$ From 2, by logic
- (4) NR $[PAST \ \supset \ (LAWS \ \supset \ PRESENT)]$ From 3, by (A)
- (5) NR (PAST) Assumption
- (6) NR (LAWS \supset PRESENT) From 4, 5, by Transfer NR
- (7) NR (LAWS) Assumption
- (8) NR (PRESENT) From 6, 7, by Transfer NR

Therefore, if determinism is true, then no one is, or ever has been, even partly morally responsible for PRESENT, that is, for any given fact that obtains now.³ Let us call this argument the “Direct Argument” or “DA” for short.⁴

DA has been welcomed by incompatibilists for a variety of reasons. Van Inwagen ([1983](#)) thought that he could, by means of it, establish the incompatibility of moral responsibility and determinism without employing the “could-have-done-otherwise” or avoidability notion of free will, and in this way sidestep disputes about the proper analysis of this notion (p. 183). Others have found DA attractive because of related worries compatibilists have voiced about the incompatibilist defense, via the Consequence Argument (van Inwagen [1983](#), pp. 93–95), of premise (1) of the Traditional Argument. Since DA does not employ the notion of avoidability, it is not subject to those worries.⁵ Finally, some incompatibilists were dissatisfied with the Traditional Argument because they became convinced by Harry Frankfurt that premise (2) of that argument (PAP) is false.⁶ These incompatibilists, however, still hold that determinism is incompatible with moral responsibility. Naturally, they too have welcomed DA, precisely because it dispenses with the notion of avoidability, a notion that they think is irrelevant to moral responsibility.⁷ Although differing in their motivation, all of these parties saw in DA an important dialectical tool that could

be used to convince compatibilists (or those who are undecided) of the incompatibility of moral responsibility and determinism.

Robert Kane, too, has argued for the incompatibility of determinism with moral responsibility. He does so on the basis of his concept of ultimate responsibility, or UR. “The basic idea is this: to be *ultimately responsible* for an action, an agent must be responsible for anything that is a sufficient reason, cause, or motive for the action’s occurring” (Kane 2005, p. 121).⁸ Based on this concept, we may put forward the following incompatibilist argument:

- (1) If determinism were true, then (given the laws of nature), *every* human act would have a sufficient cause that occurred in the distant past, prior to the existence of human beings.
- (2) But no human being is responsible for something that occurred prior to the existence of human beings.
- (3) Therefore, if determinism were true, then no human being would be ultimately responsible for any action.⁹

This argument shares with DA the dialectical advantages of using neither IVD nor PAP as a premise and of not employing *explicitly* the controversial notion of avoidability.

However, any argument based on UR also has an important dialectical disadvantage: it would not convince a compatibilist. That is, a compatibilist would claim that an agent can be fully morally responsible for an action without being morally responsible for everything that is a sufficient reason, cause, or motive for the action’s occurring; and that therefore, despite the UR-based argument, one can be morally responsible for one’s actions even if determinism is true. In fact, applying contraposition to Kane’s characterization of UR quoted above, we obtain a version of the fundamental incompatibilist intuition that if an action is a causal result of factors for which the agent is not responsible, then the agent is not responsible for the action; and of course, the compatibilist, as such, denies the validity of this intuition.

This disadvantage is not shared by DA, since it does not employ the notion of ultimate responsibility. Nevertheless, DA has recently come under severe attack from various quarters. In this essay, we offer a critical assessment of DA as an argument intended to refute compatibilism, or

at least to shift the burden of proof onto the compatibilist. We examine a variety of objections to DA raised by John Fischer and Mark Ravizza (1998), David Widerker (2002), Ishtiyaque Haji (2009), Michael McKenna (2008b), and Seth Shabo (2010). We divide these into objections based primarily on dialectical considerations (section 1) and objections in the form of counterexamples to the Transfer NR inference rule involved in DA (section 2), although dialectical considerations will sometimes be relevant to the assessment of counterexamples. In the end, we contend that the proponent of DA can deal plausibly with most of the objections that have been raised to it.

2. Dialectical Objections

2.1 McKenna's dialectical objection

McKenna (2008b, p. 370) argues that it was dialectically improper for van Inwagen to use Transfer NR in the context of the debate between compatibilists and incompatibilists, and that therefore DA should never have gotten off the ground.

To see what exactly is at issue, consider how van Inwagen justified the use of Transfer NR in DA. He did this by appealing to examples such as:

NR (John was bitten by a cobra on his thirtieth birthday)

NR (John was bitten by a cobra on his thirtieth birthday \Rightarrow John died on his thirtieth birthday)

\therefore NR (John died on his thirtieth birthday)

and

NR (Plato died in antiquity)

NR (Plato died in antiquity \Rightarrow Plato never met Hume)

\therefore NR (Plato never met Hume)¹⁰

(Following McKenna, we shall call the two examples “Snakebite” and “Plato,” respectively.)

Van Inwagen claims that the above arguments are intuitively valid and that their validity seems to be due to their exemplifying Transfer NR. Hence, these arguments show that *prima facie* Transfer NR is a valid rule of inference. Anyone who, at this point, wishes to deny the validity of Transfer NR bears the burden of proof.

But McKenna disagrees. He argues that van Inwagen's examples fail to establish Transfer NR as a rule of inference that can properly be employed in an argument against compatibilism. The reason for this failure, according to McKenna, is that neither of the examples involves a causal chain that passes through "a normally functioning agent who exercises unimpaired deliberative capacities" (McKenna [2008b](#), p. 376). Yet, DA was meant to apply precisely to cases involving such chains, since it is just such cases that are the focus of the debate between compatibilists and incompatibilists. Thus, the examples that van Inwagen adduces in support of Transfer NR are irrelevant to this debate. McKenna concludes that, with respect to the propriety of using Transfer NR in DA, the burden of proof is not on the critic to produce a counterexample to that rule; rather it is on the proponent of DA to provide supporting examples of the right sort for the rule. In the absence of support of the right kind, Transfer NR, and therefore DA, has no dialectical clout in the debate over compatibilism (McKenna [2008b](#), p. 377).

2.1.1 Response #1: Meeting McKenna's challenge

One way of responding to McKenna is to provide just the kind of example that he claims van Inwagen has failed to provide, that is, an intuitively valid instance of Transfer NR that involves a causal chain that passes through a normally functioning agent. Consider the following example, which we may call "Reprimand":

One morning, the commanding officer in the army's anti-missile defense system receives a call from one of his subordinates, Jones. Jones says he is ill and cannot come in today. Unknown either to the officer or to Jones, there happens to be a man, Smith, in the area who looks exactly like Jones. Later in the day, the officer is on his way to another base when he sees Smith, whom he takes to be Jones, frolicking on

the beach. The officer, after some deliberation, resolves to give Jones a severe reprimand the next day, which he does.

Jones, of course, does not deserve the reprimand. But is the officer morally responsible—that is, blameworthy—for reprimanding him? It seems not; for he is not blameworthy for believing that Jones was frolicking on the beach when he should have been defending the country, and he is not blameworthy for this belief’s leading him to reprimand Jones. That is, using “B” to stand for “the officer believes that Jones was frolicking on the beach” and “R” for “the officer reprimands Jones,” we can argue:

$$\text{NR(B), NR(B} \supset \text{R) } \therefore \text{NR(R).}$$

This argument obviously has the form of Transfer NR, and it seems to be intuitively valid.

Furthermore, the officer is “a normally functioning agent who exercises unimpaired deliberative capacities.” So we seem to have here just the kind of example that McKenna requires.¹¹

2.1.2 Response #2: The Dialectical Situation

Another way to respond to McKenna’s argument is to challenge his assessment of the dialectical situation—in particular, his claim that van Inwagen’s examples fail to support the validity of Transfer NR in a way that is relevant to the compatibilist/incompatibilist debate. We think that van Inwagen’s strategy is dialectically appropriate, and that by introducing Transfer NR into the debate, he has succeeded in shifting the burden of proof onto the compatibilist. In this sense, he has advanced the debate between compatibilists and incompatibilists. McKenna’s response indicates that he does not fully appreciate van Inwagen’s strategy.

To see this, let’s describe van Inwagen’s strategy in some detail. First, let’s clarify the role played by the examples Snakebite and Plato in his argument. Recall Snakebite:

NR (John was bitten by a cobra on his thirtieth birthday)

NR (John was bitten by a cobra on his thirtieth birthday \supset John died on his thirtieth birthday)

\therefore NR (John died on his thirtieth birthday)

The first thing to note is that van Inwagen was not trying to inductively generalize from it (and similar examples) to the conclusion that most likely, all instances of Transfer NR are valid. (If he were, we might wonder how he could get away with generalizing from just two examples.) Rather, the role of the examples was to elicit a certain logical and conceptual intuition. We realize, through reflecting on an example like Snakebite, that if its premises are true, the conclusion *must* be true as well, and that therefore, Transfer NR, which encapsulates this intuition, must be valid.

To see this more clearly, let's apply what we call "the reverse argument" to Snakebite. Suppose Alice claims that someone *is* responsible for John's dying on his thirtieth birthday. We ask Alice whether she thinks that someone set the snake on John, and therefore is responsible for the snake's biting John. She says: No. We then ask her whether she thinks that someone was in a position to stop the snakebite from killing John—for instance, by sucking out the venom before it entered John's bloodstream—but intentionally refrained from doing so, and therefore is (partially) responsible for the conditional fact that if the snake bites John on his birthday, then John dies on his birthday. She again answers: No. Alice's claim is very puzzling; for since she admits that no one is responsible for the snake's biting John or for the bite's leading to John's death, how can she say that someone is responsible for John's death? Our puzzlement at Alice's claims indicates that we find Snakebite, and thus Transfer NR, valid.

Once the logical and conceptual intuition expressed by Transfer NR has been elicited by examples like Snakebite and Plato, Transfer NR acquires the status of being *prima facie* valid—always.¹² For if a rule of inference, or an argument form, is valid, then it is *always* valid. With Transfer NR, van Inwagen now has at his disposal a new principle which is distinct from the principles to which incompatibilists standardly appeal. The new principle (which Transfer NR expresses in the form of a rule of inference) is that *nonresponsibility transfers across a material conditional*, or, more precisely, that nonresponsibility transfers from antecedent to consequent across a material conditional for (the truth of) which no one is responsible.¹³ At this point, van Inwagen is *prima facie* justified in applying this new principle, via DA, to all cases, including the controversial cases involving normal agency. The burden of proof is on the person who denies the unrestricted validity of Transfer NR. She must either (a) provide a clear counterexample to Transfer

NR, or else (b) provide a plausible alternative to Transfer NR, which, like Transfer NR, would account for the validity of Snakebite but which, if substituted for Transfer NR in DA, would not yield an argument against compatibilism.¹⁴

McKenna does not attempt to satisfy (a). However, one might suggest that implicit in McKenna's discussion is a proposal that satisfies (b); for he says:

If the only uncontroversial cases one can cite to establish Transfer NR are such cases [i.e., cases like Snakebite and Plato], this strongly indicates that Transfer NR is best restricted to cases that do not “pass through” [normally functioning] ...agents.

(McKenna [2008b](#), p. 376)

In other words, McKenna proposes to account for the intuitive validity of the likes of Snakebite by appeal to a restricted version of Transfer NR, excluding from its range of application cases of normal agency. Note, however, that those excluded cases are precisely the controversial cases over which the incompatibilist and the compatibilist disagree. McKenna does not explain *why* Transfer NR should not apply to these cases; he simply asserts, without explanation, that the compatibilist is entitled to claim that it does not apply. So it seems that what McKenna recommends, in effect, is that Transfer NR is valid, except where it conflicts with compatibilist intuitions. To us, McKenna's suggestion seems ad hoc and dialectically improper.¹⁵

2.2 Widerker's Dialectical Criticisms of DA

Widerker has raised two dialectical objections against DA.

(W1) DA depends for its soundness on the assumption that determinism is incompatible with avoidability (IVD), an assumption which, since it involves the contested notion of avoidability, is controversial between compatibilists and incompatibilists. Hence, DA does not have the dialectical force envisaged for it by van Inwagen.¹⁶

(W2) The rationale for accepting the premise NR(LAWS) of the Direct Argument is that we do not have the power to render the laws of nature false, which means that

the argument tacitly assumes the unwanted notion of avoidability that van Inwagen thought he could do without (Widerker [2002](#), p. 324).

2.2.1 Assessment of (W1)

To establish (W1), Widerker describes a deterministic scenario in which an agent, Jones, deliberately performs a heinous act A for a selfish reason, knowing very well that in doing so he is acting immorally, and believing that he could have done otherwise. Widerker then raises the question whether Jones is blameworthy for A. Incompatibilists would want to answer this question negatively, since the scenario is deterministic. But suppose, Widerker continues, someone claims that Jones *could have avoided* acting as he did, and that therefore Jones *is* blameworthy; after all, he knew that he was acting wrongly and knew also that he could have acted otherwise. At this point, the proponent of DA seems to be faced with a dilemma. He cannot reply that, since the scenario under consideration is deterministic, Jones *could not* have refrained from acting immorally, for this would be, in effect, assuming (IVD)—something the proponent of DA would not want to do. On the other hand, he cannot afford to ignore the question of the avoidability of Jones's act. Doing so would mean tacitly legitimizing the claim that Jones could have avoided acting immorally *despite* determinism, and therefore is blameworthy for what he did; and this, in turn, would mean that DA is unsound. Widerker also points out that the defender of DA cannot evade this objection by arguing that he is more sure of the soundness of DA than he is of an assumption that the objection presupposes, that is,

(M) Someone who knowingly acted in a morally wrong way, and believed correctly that he could have avoided acting as he did, is blameworthy for his act.

For this assumption is beyond reproach, being (almost) self-evident.¹⁷ The conclusion Widerker ([2002](#)) draws from the above considerations is that one cannot maintain the soundness of DA without assuming the truth of (IVD), and therefore, DA cannot do the work van Inwagen wanted it to do (pp. 322–323).

Some have claimed that Widerker's objection (W1) does not succeed, since assumption (M), upon which it is based, is *not* beyond reproach, but is in fact false. McKenna ([2008b](#)), for instance, offers a counterexample to (M) in which an agent satisfies the conditions that (M) specifies, but is not morally responsible for her action because she is mentally deranged (p. 358). However, his counterexample is based on a misunderstanding of Widerker's position. (M), as intended by Widerker, was meant to apply to a morally competent, normally functioning agent, whose mental health is not in question. In other words, (M) is to be understood as having an implicit clause to the effect that "all else is normal."¹⁸

Though this critique of Widerker's (W1) fails, we would now like to argue that (W1) is not nearly as damaging to DA as Widerker originally thought. Widerker wanted to show that the proponent of DA must *assume* (IVD), contrary to the main purpose of DA. However, all that follows from (W1) is that accepting DA *commits* one to holding (IVD). That is, it does *not* follow that DA *presupposes* (IVD); rather (IVD) is a logical *consequence* of DA plus (M). For given that the scenario that Widerker employs is a deterministic one, DA entails that in that scenario Jones is not blameworthy for what he did; and, given that he knowingly and deliberately acted in a morally wrong way, and also believed that he could have avoided acting as he did, it follows, via (M), that in that scenario he could *not* really have avoided acting wrongly.

Widerker's argument (W1) *does* entail that the proponent of DA must assume *something* about avoidability—namely (M). But whereas (IVD) is controversial, (M) is quite harmless and would be accepted by both compatibilists and incompatibilists. This point is important since it enables incompatibilists to restate the dialectical role of DA in a way that evades Widerker's worry. In lieu of viewing it as an argument that does not employ the avoidability notion of free will at all, incompatibilists may view DA as an argument that is free of any *controversial* assumptions involving avoidability. The case for the incompatibility of determinism and moral responsibility based on DA would remain unimpaired.

2.2.2 Assessment of (W2)

We turn now to (W2), Widerker's second dialectical objection to DA. An essential premise of DA is that no one is responsible for the laws of nature—NR(L). But why, asks Widerker (2002) rhetorically, should one accept that premise, if not for the fact that no one could have prevented them from obtaining (p. 324)? That is, Widerker argues that

(W-L) Unless one assumes that no one could have prevented the laws of nature from obtaining, one has no good reason to accept NR(L).

If Widerker is right, then DA depends for its plausibility on an assumption involving unavoidability, contrary to van Inwagen's motivation for introducing DA.

Both McKenna and Haji reject Widerker's contention (W-L). They propose several ways of justifying acceptance of NR(L) that are not based on the laws' unavoidability:

- (1) No one is responsible for the laws of nature because no human being brought them about;
- (2) No one is responsible for the laws of nature because no human being *could* bring them about;
- (3) One is not responsible for the laws of nature because one did not bring them about, and they obtain regardless of whether one exists or fails to exist;

and, adopting a proposal made by Ted Warfield (1996, p. 218),

- (1) No human being is responsible for the laws of nature because their *truth-makers* operated before human beings came into existence.

Hence, they claim that one can account for our nonresponsibility for the laws of nature, without assuming their unavoidability (McKenna 2008b, p. 360; Haji 2009, pp. 84–85).

However, we do not think that (1)–(4) either singly or jointly account adequately for our nonresponsibility for the laws of nature. Consider the following example, which we may call “The Miracle Worker”:

Suppose that the reasons for the nonresponsibility for the laws of nature stated in (1)–(4) all obtain; that is, no human being brought about, or could have brought about, the laws of nature; the laws hold regardless of whether any particular person exists; and the truth makers of the laws of nature obtained before any human being

was born. But suppose that nevertheless, someone named Smith somehow has the power to bring it about that a given law of nature L1 does not obtain in a particular instance, and knows that he has it. Suppose further that Smith knows that by exercising this power, he could prevent a tragedy from occurring; but he intentionally refrains from suspending L1, for morally inadequate reasons, and the tragedy occurs.

Note that if (1)–(4) were adequate explanations of our nonresponsibility for the laws of nature, as McKenna and Haji maintain, then Smith would not be responsible for the obtaining of L1 in this scenario, and therefore he would not be responsible for the tragedy. However, it is clear that Smith *is* responsible (blameworthy) for the tragedy, and part of the reason why Smith is responsible for the tragedy is that he is responsible for L1's obtaining in the circumstances; for Smith knew that L1's obtaining would lead to the tragedy, and he knew that he could prevent L1's obtaining (in the circumstances), and yet he did not prevent it. This example also shows that what renders (1)–(4) inadequate as explanations of our nonresponsibility for the laws of nature is precisely their failure to take into account the *unavoidability* of the laws.

Note that one cannot parry this objection by arguing that no human being could possibly have it within his power to prevent a law of nature from obtaining, and that therefore the objection rests on an assumption that is necessarily false. If this is the critic's only objection to our example, then he is *ipso facto* conceding Widerker's claim (W-L) that to be able fully to account for our nonresponsibility for the laws of nature we have to assume their unavoidability.

Thus Widerker's claim (W2) seems to be correct, which means that strictly speaking, van Inwagen has not achieved his goal of constructing an argument for incompatibilism without appealing to assumptions involving avoidability. However, having said that, we must confess that this particular assumption—that is, the unavoidability of LAWS—strikes us as so intuitively plausible that it would be fair to say that DA's dependence on it would not detract significantly from its philosophical importance.¹⁹ As emphasized earlier, DA's dialectical role can be easily reformulated, that is, DA may be viewed as an argument for the incompatibility of moral responsibility and determinism that is free of any *controversial* assumptions about avoidability.

3. Alleged Counterexamples to Transfer NR

We turn now to a critical discussion of proposed counterexamples to Transfer NR.

3.1. Ravizza and Fischer's Counterexample

Mark Ravizza and John Fischer have presented the following a scenario, known as “Erosion”:

Betty [a double agent who has been instructed to destroy the enemy's camp] plants her explosives in the crevices of a glacier and detonates the charge at $T1$, causing an avalanche that crushes the enemy's fortress at $T3$. Unbeknownst to Betty...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 at $T2$, starting a natural avalanche that would have crushed the enemy's camp at $T3$. (Ravizza [1994](#), pp. 72–73; Fischer and Ravizza [1998](#), p. 157)

They extract from the above scenario the following argument:

NR (the glacier was eroding)

NR (the glacier was eroding \supset the enemy's camp was destroyed at $T3$)

\therefore NR (the enemy's camp was destroyed at $T3$)

According to the scenario, the premises are true, but the conclusion is false. Since the argument has the form of Transfer NR, we see that Transfer NR is not valid.

3.1.1 McKenna's Response

Michael McKenna noted that this example is based on there being two causal paths to the same result—that is, either the glacier's eroding or Betty's explosives would result in the destruction of the enemy camp. So McKenna ([2008b](#)) suggested a version of Transfer NR that rules out such “two-track” cases by stipulating that there be just one causal path between p and q. His new principle is:

Transfer NR^{***}: If a person is not responsible for a fact, and this fact is at some particular time the sole causal antecedent of the path to the bringing about of a further fact, and if the person is not responsible for any part of the path between the two facts, then he or she is not responsible for the further fact either. (p. 368)²⁰

Fischer has responded that this “single-track” version of the transfer rule is question-begging. We think that what Fischer means is that McKenna’s version essentially involves reference to causality in a way that makes it merely a restatement of the standard incompatibilist intuition, which compatibilists reject, namely that no one is responsible for what is *causally determined* by factors for which no one is responsible. As we pointed out earlier, Transfer NR, unlike the standard incompatibilist intuition, is supposed to be acceptable to an open-minded compatibilist. So McKenna, by replacing Transfer NR with Transfer NR^{***}, which is essentially a statement of the standard incompatibilist intuition, has taken a dialectical step backwards.²¹

3.1.2 Widerker’s Response

David Widerker (2002, pp. 318–319) has argued that Fischer and Ravizza’s example can be sidestepped by drawing attention to the fact that its conclusion involves *derivative* responsibility. An agent S is *derivatively* responsible for p just in case S is responsible for p in virtue of being responsible for some other fact. S is *directly responsible* for p just in case S is responsible for p, but *not* in virtue of being responsible for some other fact. In Erosion, Betty is derivatively responsible for the destruction of the enemy’s camp by virtue of being responsible for detonating the charge that caused the destruction. The notion of derivative responsibility is parasitic upon the notion of *direct* or *nonderivative* responsibility, and is therefore only of secondary importance in the debate between compatibilists and incompatibilists.²² In other words, necessarily, if no one is directly responsible for anything, then no one is derivatively responsible for anything. Thus, if one were able to show that determinism is incompatible with direct responsibility, it would immediately follow that it is incompatible with derivative responsibility as well.²³

3.2 Haji’s Example

A very interesting counterexample to the validity of Transfer NR has been offered by Ishtiyaque Haji (2009, p. 97). He calls it “Hal’s Creation”:

Hal is an amoral creator who creates a world W_1 , which is indeterministic. In W_1 , Michael, an agent with libertarian free will, kindly (and freely) decides at time T to help his friend Bill.

Given this scenario, Haji claims that the following Transfer NR argument has true premises and a false conclusion:

1H. NR (Hal creates W_1)

2H. NR (Hal creates $W_1 \Rightarrow$ Michael decides at T to help Bill)

\therefore 3H. NR (Michael decides at T to help Bill)

1H is true because Hal is the only candidate for moral responsibility here, and he is stipulated to be amoral. 3H is false because Michael is praiseworthy for deciding to help his friend. Haji’s justification of 2H is more subtle and complex. First, he puts forward two alternative epistemic constraints on moral responsibility:

(E1) To be morally praiseworthy [blameworthy] for bringing about a state of affairs p , the agent must have brought it about in the belief that p is a morally good [morally bad] state of affairs.

(E2) To be praiseworthy [blameworthy] for some state of affairs p , the agent *ought to have known or believed* that p is a morally good [morally bad] state of affairs.

(Haji 2009, p. 99)²⁴

Haji argues that neither Hal nor Michael can be said to satisfy either of these constraints with respect to the conditional state of affairs MC:

(MC) Hal creates $W_1 \Rightarrow$ Michael decides to help Bill at T .

This is obviously true for Hal, since Hal is amoral; but Haji thinks that it is also true for Michael, since there is no reason to think that Michael brought about MC in the belief that MC is a good state of affairs, nor is it true that he should have believed that MC is a morally good state of

affairs. Finally, Haji points out (or stipulates) that if Hal and Michael are not responsible for MC, then no one is. Hence, he claims that 2H is true, and so Hal's Creation is a counterexample to Transfer NR.²⁵

3.2.1 Our Response

Unlike Haji, we think that the argument based on Hal's Creation is *valid*. We agree with Haji that the conclusion of the argument (3H) is false, that is, Michael *is* responsible (praiseworthy) for deciding to help Bill. But we would argue that its second premise

2H. NR (Hal creates W1 \supset Michael decides at T to help Bill).

is false as well. To see this, note that MC (the state of affairs which 2H says that no one is responsible for) is the same state of affairs as the disjunction

(MC') \sim (Hal creates W1) \vee (Michael decides to help Bill at T),

which, given the falsity of its first disjunct, obtains *just* in virtue of Michael's deciding to help Bill.

But then, since Michael is praiseworthy for deciding to help Bill, he can be said to be *derivatively* responsible for MC' (and hence for MC), in virtue of being responsible for his decision to help Bill.²⁶ Hence, the second premise of Haji's argument is false.

As for Haji's epistemic constraints on responsibility to which he appeals in order to defend 2H, we would argue that these do not always apply to derivative responsibility, the kind of responsibility involved in (2H). For example, Jones might be praiseworthy for defusing a bomb planted in a mall, and hence be (derivatively) praiseworthy for the state of affairs consisting in little Jimmy's life being saved. Yet, since he had no clue as to Jimmy's existence or identity, he did not bring about that state of affairs in the belief that Jimmy's life being saved is a good state of affairs, nor *ought* he to have believed that Jimmy's life being saved is a good state of affairs (since, again, he knew nothing of Jimmy's existence).

3.3 Shabo's Counterexample

Seth Shabo (2010) has also put forward a counterexample to the validity of Transfer NR. He bases it on a scenario which he calls “Bad Angle”:

Jed murders Kenny by dislodging a boulder so that it falls onto him. A ranger could have prevented the murder by warning Kenny; but since he was viewing things from a bad angle, he misperceived the situation so that he mistakenly believed that Kenny was not in the boulder’s path, and so he did not warn him. (pp. 243–244)

The counterexample that Shabo extracts from this scenario is rather complex. Here is a simplified version which, we believe, does not detract from its main thrust. Let “M” stand for “The ranger misperceives the situation so that he mistakenly believes that the boulder will not hit Kenny,” and let “K” stand for “The boulder hits Kenny.” Consider now the following instance of Transfer NR:

$$\text{NR}(M), \text{NR}(M \supset K) / \therefore \text{NR}(K)$$

This argument has true premises and a false conclusion, which implies that Transfer NR is invalid. The first premise is true, since no one is responsible for the ranger’s misperception. The second premise is true, since $(M \supset K)$ is a necessary truth, and so no one is responsible for its truth. But the conclusion is false, since Jed is blameworthy for K.

However, we think that the first premise, $\text{NR}(M)$, is false as well. Note that M is true partially in virtue of the ranger’s mistakenly believing that the boulder will *not* hit Kenny.²⁷ But mistakenly to believe that the boulder will not hit Kenny means to believe that the boulder will not hit Kenny, when in fact the boulder *will* hit Kenny. Thus, M is true partially in virtue of its being the case that the boulder will hit Kenny. But since Jed is responsible for the boulder hitting Kenny, it follows that Jed is partially responsible for M. Hence it is false that $\text{NR}(M)$, and therefore we do not have here a counterexample to Transfer NR.

3.4 A New Example: Molecules

Suppose that Jones fires a bullet at Smith at T1, a fact which, in the circumstances, provides a causally necessary and sufficient condition for the movement of some air molecules at T3 in his vicinity. The following argument shows that Transfer NR is invalid.

1. NR (Some air molecules move at $T3$) (True)
2. NR (Some air molecules move at $T3$ \supset Jones fires a bullet at Smith at $T1$) (True)
- \therefore 3. NR (Jones fires a bullet at Smith at $T1$) (False)²⁸

Premise 1 is true, since intuitively the fact that some air molecules are moving at $T3$ is not something for which anyone is *morally* responsible, in these circumstances. And since we assumed that Jones's firing a bullet at Smith at $T1$ is a causally necessary condition of the fact that some air molecules are moving at $T3$, premise 2 is true as well. However, the conclusion is false, since Jones is responsible for firing a bullet at Smith.

This counterexample might be blocked by a further modification of Transfer NR. We might restrict Transfer NR to cases in which it is not the case that q *temporally precedes* p .²⁹ Some might find this restriction ad hoc. Thus, dealing with Molecules requires a change in Transfer NR, which *may* detract from the rule's broad intuitive appeal to some extent. We leave it to the reader to evaluate this worry.³⁰

4. Conclusion

The picture that emerges on the basis of our assessment of DA is the following. Of the dialectical objections to DA, we have seen that McKenna's is unsuccessful, whereas Widerker's objections, though successful, nevertheless are not as damaging to DA as they initially seemed to be. As for the alleged counterexamples to Transfer NR, they are either unsuccessful or they can be adequately dealt with by relatively minor alterations in Transfer NR. Thus, DA seems to emerge from its battles somewhat weakened, but still a powerful argument that advances the compatibilist/incompatibilist debate by shifting the burden of proof onto the compatibilist.³¹

References

- Campbell J.K. 2006. "Farewell to Direct Source Incompatibilism." *Acta Analytica* 21: 50-61.
- Fischer, J.M. and Ravizza, M. 1998. *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.
- Fischer, J.M. 2006. "The Transfer of Nonresponsibility." In *My Way: Essays on Moral Responsibility*. Oxford: Oxford University Press, 159-174.
- Flint, T. 1987. "Compatibilism and the Argument from Unavoidability." *Journal of Philosophy* 84: 423-40.
- Frankfurt, H. 1969. "Alternate Possibilities and Moral Responsibility." *Journal of Philosophy* 66: 829-39.
- Ginet, C. 2003. "Libertarianism." In *The Oxford Handbook of Metaphysics*, ed. Michael J. Loux and Dean W. Zimmerman. New York: Oxford University Press.
- Ginet, C. 2007. "An Action Can be Both Uncaused and Up to the Agent." In *Intentionality, Deliberation, and Autonomy*, ed. Christop Lumer and Sandro Nannini. Aldershot, UK: Ashgate, 243-255.
- Haji, I. 2009. *Incompatibilism's Allure*. Peterborough: Broadview Press.
- Horgan, T. 1985. "Compatibilism and the Consequence Argument." *Philosophical Studies* 47: 339-56.
- Hunt, T. 2000. "Moral Responsibility and Avoidable Action." *Philosophical Studies* 97: 195-227.
- Kane, R. 2005. *A Contemporary Introduction to Free Will*. Oxford University Press.
- Kane, R. 1996. *The Significance of Free Will*. Oxford University Press.
- Kapitan, T. 2002. "A Master Argument for Incompatibilism?" In *The Oxford Handbook of Free Will*, ed. Robert Kane. Oxford: Oxford University Press, 127-157.
- Lewis, David. 1981. "Are We Free to Break the Laws?" *Theoria* 47: 113-121.
- McKenna, M. 2001. "Source Incompatibilism, Ultimacy, and the Transfer of Non-responsibility." *American Philosophical Quarterly* 38: 37-51.
- McKenna, M. 2008. "Saying Good-bye to the Direct Argument the Right Way." *Philosophical Review* 117: 349-383.
- Pereboom, D. 2001. *Living without Free Will*. Cambridge: Cambridge University Press.
- Ravizza, M. 1994. "Semicompatibilism and the Transfer of Non-Responsibility." *Philosophical Studies* 75: 61-94.
- Rowe, W. L. 1980. "Fatalism and Truth." *Southern Journal of Philosophy* 18: 213-219.
- Schnall, I. M. and Widerker, D. 2012. "The Direct Argument and the Burden of Proof." *Analysis* 72: 25-36.
- Shabo, S. 2010. "Against Logical Versions of the Direct Argument: A New Counterexample."

American Philosophical Quarterly. 47: 239-252.

Stump, E. 1999. "Alternative Possibilities and Moral Responsibility: The Flicker of Freedom." *The Journal of Ethics* 3: 299-324.

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

Warfield, Ted. 1996. "Determinism and Moral Responsibility Are Incompatible." *Philosophical Topics* 24: 215–26.

Widerker, D. 2002. "Farewell to the Direct Argument." *Journal of Philosophy* 99: 316–24.

Widerker, D. 2008. "Some Further Thoughts on the Direct Argument." In *Essays on Free Will and Moral Responsibility*, ed. Nick Trakakis and Daniel Cohen. Newcastle: Cambridge Scholars Press. 224-233.

Zagzebski, L. 2000. "Does Libertarian Freedom Require Alternate Possibilities?" *Philosophical Perspectives* 14: 231–48.

¹ An explanation of the “even partly” phrase is in order. Van Inwagen ([1983](#)) inserted it in order to forestall a certain type of counterexample to Rule B. For example, consider the following argument:

NR(John throws a six)

NR(John throws a six \supset John plays dice)

Therefore, NR(John plays dice).

If we omit the “even partly” phrase from the definition of “NR(p),” then it is easy to imagine a situation in which the premises are true but the conclusion false, and we have a counterexample to Rule B. However, if “even partly” is included, then, in the situation under consideration, the premise “NR(John throws a six)” is false, since John is *partly* responsible for throwing the six, by virtue of being responsible for playing dice in the first place, and so this argument is not a counterexample to Rule B (p. 243, n. 28). In that footnote, Van Inwagen also suggests that, instead of including the “not even partly” in the definition of “NR(p),” one could define “NR(p)” as “p, and no human being is morally responsible for p *or for any logical consequence of p.*” However, it is far from clear that this latter definition fully captures the intuitive meaning of “not being even partly responsible for p.” Hence, in what follows, we prefer to use van Inwagen’s original definition of “NR(p),” relying on an intuitive understanding of “being partly responsible” as bearing *some* responsibility for the fact that p. On this point, see Shabo ([2010](#), pp. 244–246).

² Van Inwagen uses “N” where we use “NR.”

³ From now on, for simplicity, we will use “responsible” to mean “at least partly responsible” (and “not responsible” to mean “not even partly responsible”), unless otherwise indicated.

⁴ We borrow the term “Direct Argument” from John Fischer and Mark Ravizza ([1998](#), ch. 6).

⁵ For those compatibilist worries, see, for example, Lewis ([1981](#)), Horgan ([1985](#)), Flint ([1987](#)), and Kapitan ([2002](#)).

⁶ See Frankfurt ([1969](#), pp. 829–839).

⁷ Here we have in mind, for example, Stump (1999), Zagzebski (2000), Hunt (2000), and Pereboom (2001, p. 33). These incompatibilists are sometimes referred to as “Causal History Incompatibilists” or as “Source Incompatibilists.”

⁸ For a more elaborate definition of this concept, see Kane (1996, p. 35).

⁹ Our UR-based argument differs somewhat from Kane’s own. See, for example, Kane (2005, pp. 122–123). We prefer to use ours simply because of its brevity. The remarks in the text on our argument apply also to Kane’s.

¹⁰ The Plato example is from van Inwagen (1983, p. 187).

¹¹ For a more elaborate discussion of this example, along with another one, see Schnall and Widerker (2012, pp. 30–31). In particular, we argue there that anyone who denies the conclusion of the Reprimand-argument is committed to denying one of the premises—a strong indication of its validity.

¹² This status can be challenged—for instance, by presenting a counterexample to the validity of that rule. However, as long as such an example has not been presented, and as long as no alternative way to explain the intuition of validity is forthcoming, we are justified in regarding Transfer NR as a valid rule of inference.

¹³ Note that this is not the same as the principle that nonresponsibility is preserved through causal determination by factors for which the agent was not responsible.

¹⁴ For a beautiful example of how a philosophical principle may be disarmed by explaining away the examples that are used to support it, see William Rowe’s (1980) refutation of Richard Taylor’s argument for fatalism.

¹⁵ For a more extended discussion of these dialectical issues, see Schnall and Widerker (2012, pp. 31–35).

¹⁶ Widerker (2002, pp. 322–323) originally aimed this argument at the validity of Transfer NR or any improved version thereof. However, we think that the argument is best construed as targeting DA in general. Of course, one who views Transfer NR as the most (or the only) vulnerable aspect of DA will interpret the argument as implicating Transfer NR specifically.

¹⁷ Note that (M) does not assume that avoidability is a necessary condition for moral responsibility (PAP). What it assumes is only that avoidability in conjunction with certain other assumptions is sufficient for moral responsibility.

-
- ¹⁸ We also think that the counterexample itself is problematic, as it is far from clear that the agent in the example, given that she is stipulated to be morally competent (despite being deranged), is not blameworthy for her action.
- ¹⁹ That said, note that there are compatibilists who do deny this assumption. Here we have in mind Altered-Law Compatibilists such as Lewis ([1981](#)).
- ²⁰ We skip over an earlier version of McKenna's ([2001](#)) single-track principle (p. 45). Fischer's response, in the text below, was to the earlier version, but it applies equally to this one. Note also that although McKenna formulates Transfer NR*** as a conditional, we prefer to treat it as a rule of inference, since it is supposed to be a substitute for Transfer NR. The formal changes involved do not affect the discussion to follow.
- ²¹ McKenna attempts to avoid the charge of begging the question by construing the causality involved in Transfer NR*** as probabilistic. However, this attempt is unsuccessful; for interpreting Transfer NR*** this way renders that principle false. The fact that one's act is (merely) probabilistically caused by factors for which one is not responsible does not take one off the hook.
- ²² Similar remarks would apply to Widerker's Brain Malfunction argument in Widerker (2008b<AU: Widerker (2008b) has not been provided in the Bibliography. Please check.>, p. 231), the conclusion of which also involves derivative responsibility.
- ²³ Working out this response in greater detail would require, among other things, limiting the conclusion of Transfer NR to absence of direct responsibility rather than nonresponsibility *sans phrase*. We believe that the relevant details can be worked out.
- ²⁴ Haji's own formulation of these constraints differs from that in the text, but this difference can be safely ignored.
- ²⁵ Haji ([2009](#)) also offers another counterexample to Transfer NR for the case that (W1) is a deterministic world, meant to appeal specifically to compatibilists (p. 98). However, Haji's argument strikes us as overly dogmatic, as it simply has the compatibilist reject the Transfer NR-rule without addressing van Inwagen's argument in support of its validity.
- ²⁶ Note that our claim that Michael is derivatively responsible for MC', does not commit us to the acceptance of the principle that for any propositions p and q:

If one is morally responsible for q , and p is false, then one is derivatively responsible for its being the case that $\sim p$ or q .

This principle is obviously false. It would entail that one can be responsible for necessary truths such as $(p$ or $\sim p)$. We do, however, subscribe to a weaker principle which is like the one above but constrained by the condition that $(\sim p$ or $q)$ is not an instance of a logical truth, nor a nomologically necessary truth, nor a truth that is nomologically necessitated by the circumstances.

²⁷ We may view M as equivalent to the following conjunction: (1) The ranger misperceives the situation, and (2) as a result, the ranger comes to believe that the boulder will not hit Kenny, but (3) the boulder *will* hit Kenny.

²⁸ This example is a variation on an example by Widerker ([2002](#), p. 319).

²⁹ For a similar restriction, see Ginet ([2003](#), p. 607).

³⁰ For another attempt to refute Transfer NR, which is based on a Frankfurt-type counterexample to PAP, see Campbell ([2006](#), pp. 43–44). However, since Frankfurt-cases are controversial, a full discussion of the issues involved is beyond the scope of this paper.

³¹ We would like to thank Ori Beck, Carl Ginet, Dovid Gottlieb, and Michael Pauen for excellent comments and discussions on earlier versions of this paper. <AU: This type of note is usually set as the first, unnumbered note. OK to do that here?>