The Direct Argument and the burden of proof

IRA M. SCHNALL & DAVID WIDERKER

1. Introduction

To show that determinism is incompatible with moral responsibility¹, incompatibilists have traditionally employed the following argument (henceforth the 'Traditional Argument'):

- 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: 182-88) proposed another argument for the same conclusion, one that argues 'directly' for that conclusion from some general and allegedly uncontroversial assumptions about moral responsibility and determinism.

Let 'PAST' stand for a true proposition describing the state of the world at some time before the existence of human beings, '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) $p \vdash NR(p)$
- (B) NR(p), $NR(p \supset q) \vdash NR(q)$,

¹ By 'moral responsibility' we shall understand here blameworthiness or praiseworthiness. These two notions may not exhaust the concept of moral responsibility; for one may also be morally responsible for a morally neutral act. We have two reasons for focusing on blameworthiness or praiseworthiness. First, the debate over the compatibility of determinism and moral responsibility is essentially over whether agents in a deterministic world can be blameworthy or praiseworthy. Second, establishing incompatibilism for *just* one of these notions would be a significant accomplishment.

where ' 'represents broadly logical necessity, '⊢' stands for 'therefore', '⊃' stands for material implication and 'NR(p)' abbreviates 'p and no one is (now), or ever has been even partly morally responsible for the fact that p'. 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. $[(LAWS \& PAST) \supset PRESENT]$ From 1

3. $[LAWS \supset (PAST \supset PRESENT)]$ From 2, by logic

4. NR [LAWS \supset (PAST \supset PRESENT)] From 3, by Rule (A)

5. NR (LAWS) Assumption

6. NR (PAST \supset PRESENT) From 4, 5, by Transfer NR

7. NR (PAST) Assumption

8. NR (PRESENT) From 6, 7, by Transfer NR

Therefore, if determinism is true, then no one is, or ever has been, morally responsible for PRESENT, that is, for any given fact that obtains now. Let us call this argument, following Fischer and Ravizza (1998, chapter 6), the 'Direct Argument,' or 'DA' for short.

DA seems to have several dialectical advantages over the Traditional Argument. In particular, DA seems to establish its incompatibilist conclusion without depending on either of the premises of the Traditional Argument. This is an advantage, because both premises are controversial. Compatibilists typically deny the first premise (IVD); and many philosophers on both sides of the debate deny the second premise (PAP). However, perhaps partly because DA seems to be such a simple yet powerful argument, it has come under attack, in the form of proposed counterexamples, and in the form of arguments to the effect that DA does not really succeed in avoiding controversial assumptions.³

² In lieu of 'NR', van Inwagen uses 'N'. In what follows, we leave out the 'even partly' for simplicity; this does not affect our argument.

³ See Fischer and Ravizza (1998), 157; Widerker (2002); Haji (2008).

In a recent article, Michael McKenna (2008) has mounted a novel attack on DA: He contests the propriety of van Inwagen's using the Transfer NR rule in an argument against compatibilism.

McKenna's attack raises important issues about philosophical argumentation. In this paper, we address those issues and contend that McKenna's argument does not succeed.

2. McKenna versus van Inwagen

As we said, the target of McKenna's critique of DA is the inference rule Transfer NR:

$$NR(p)$$
, $NR(p \supset q) \vdash NR(q)$.

McKenna (2008: 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, let us see how van Inwagen justified the use of Transfer NR in DA. He argued for its validity by appealing to examples such as the following (1983: 187):

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

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

... NR (John died on his thirtieth birthday);

and

NR (Plato died in antiquity);

NR (Plato died in antiquity ⊃ Plato never met Hume);

∴ NR (Plato never met Hume).

(Following McKenna, we shall call the two examples 'Snakebite' and 'Plato', respectively). According to van Inwagen, once the validity of Transfer NR is *prima facie* established by such examples, anyone who claims that it is not valid bears the burden of proof. This means, according to van Inwagen, that he will have to do one of the following two things:

[1] he will have to produce some set of propositions intuitively more plausible than the validity of [Transfer NR] and show that these propositions entail the compatibility of moral responsibility and determinism, or else [2] he will have devise a counterexample to [Transfer NR], a counterexample that can be evaluated independently of the question of whether moral responsibility and determinism are compatible.⁴ (van Inwagen 1983: 188)

2.1 McKenna's dialectical objection

McKenna claims that there is a third way to contest the use of Transfer NR in DA, and this is to argue 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 in the examples adduced by van Inwagen,

the chains of causal sufficiency through which ... nonresponsibility is transmitted never 'pass through' a normally functioning agent who exercises unimpaired deliberative capacities in the production of an (allegedly) free action for which he or she is morally responsible. (McKenna 2008: 376)

Briefly put, McKenna's point is that van Inwagen's examples do not involve *normal agency*. But DA was meant to apply precisely to cases involving normal agency; for these are just the cases about which compatibilists and incompatibilists disagree. Thus, the examples that van Inwagen adduces in support of Transfer NR are irrelevant to the debate between incompatibilists and compatibilists. McKenna (2008: 377) 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.

⁴ We think that van Inwagen's first option for his critic is a bit too strong. We would rather say that the alternative to providing a counterexample to Transfer NR is to provide a principle that does the work of Transfer NR but which is *consistent* with compatibilism. This observation, however, does not affect the discussion to follow.

2.2 McKenna's example and his analysis thereof

In order to illustrate his point, McKenna offers the following scenario: In a deterministic world, George Bush has promised his daughter that he would 'get down' (that is, that he would dance with abandon) in order to show that he is 'hip.' He deliberates whether to get down; on the one hand, if he does get down, he will be a laughing stock on nighttime TV, but on the other hand, he promised his daughter that he would get down, and his getting down would make her happy. In the end, he gets down. McKenna uses the following abbreviations:

P = a description of the state of the world at a time in the distant past

L =the laws of nature

g = a description of a chain of events starting with P and ending in George Bush's existing

r = the statement that George Bush deliberates as to whether to get down

s =the statement that George Bush gets down

McKenna then applies DA to Bush's getting down (that is, he substitutes 's' for 'PRESENT' in DA).

Since we are talking about a deterministic world, we have

D:
$$\Box$$
 (P&L. \supset s).

Starting with (D), we can run DA to get the incompatibilist conclusion NR(s). However, according to McKenna, there are problems with this application of DA. Since between P and s there is a causal chain including links g and r, it follows that if (D) is true, then so is the following:

DBD:
$$\Box$$
 (P&L. \supset g) & \Box (g&L. \supset r) & \Box (r&L. \supset s).

Applying Rule A to each of the conjuncts of (DBD), we get

DBD':
$$NR(P\&L. \supset g) \& NR(g\&L. \supset r) \& NR(r\&L. \supset s)$$
.

On the basis of (DBD'), we can construct the following argument, or chain of arguments, which McKenna calls 'Determined Bush Deliberates' (2008: 375):

- 1. NR (P & L) given
- 2. NR (P & L. \supset g) (DBD')
- 3. NRg 1, 2, Transfer NR
- 4. NR (g & L) 1, 3
- 5. NR (g & L. \supset r) (DBD')
- 6. NRr 4, 5, Transfer NR
- 7. NR (r & L) 1, 6
- 8. NR (r & L. \supset s) (DBD')
- 9. NRs 7, 8, Transfer NR

He claims that our earlier application of DA to s is sound only if the above chain of arguments in its entirety is sound; 'if one link in the chain of arguments is subject to credible doubt,' then that impugns the deduction of NR(s) from (D). Now the last argument in the chain, steps 7-9, is precisely the kind of argument about which compatibilists and incompatibilists disagree. Thus this last link in the chain is subject to credible doubt, and therefore the use of DA to prove that NR(s) is impugned. (McKenna 2008: 375)

But why not consider Transfer NR to have been established by examples like Snakebite and Plato, and then accept steps 7-9 as valid? The answer to this question is precisely McKenna's point: Examples like Snakebite and Plato, which seem to support Transfer NR, do not involve normal agency; so it is open to a compatibilist to deny that applications of Transfer NR to cases that do involve normal agency are valid. In other words, in a case where a causal path passes from P, through deliberation and decision, to an action A, the compatibilist is entitled to deny that NR(P) and NR(P \supset A) jointly entail NR(A); that is, she is entitled to deny the validity of Transfer NR in such cases – unless the proponent of DA can provide *suitable* examples in support of Transfer NR. (McKenna 2008: 376)

3. 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. Unbeknown 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 is extremely insulted by the reprimand. He denies having been on the beach, but to no avail; the officer insists 'I saw you there with my own eyes.'

Jones, of course, does not deserve the reprimand. But is the officer morally responsible – in particular, 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:

$$NR(B)$$
, $NR(B \supset R) / :: NR(R)$.

This argument is obviously of the form 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.

In order to see more clearly that this Transfer NR argument is valid, consider the following application of what we call 'the reverse argument': Suppose someone, Alice, tells us that she thinks that the officer *is* blameworthy for reprimanding Jones. We might ask Alice whether she thinks that the officer should not have jumped to the conclusion that the man he saw on the beach was Jones, and

therefore he is blameworthy for believing that it was Jones on the beach. Suppose she says 'No'; after all, the man he saw on the beach looked exactly like Jones. We might then ask whether she thinks that the officer is blameworthy for letting this belief lead him to reprimand Jones; perhaps he should have, instead, decided to speak to Jones in a more sympathetic way. Suppose she again answers 'No'; after all, if he believes that Jones was frolicking on the beach when he should have been defending his country, and that he must have been lying when he called in sick, then it is perfectly reasonable for the officer to conclude that Jones should be reprimanded. At this point, we will be very puzzled by Alice's position; for it seems that in this scenario, if the officer is blameworthy for reprimanding Jones [so \sim NR(R)], that can only be either because he is blameworthy for believing that Jones was on the beach [so \sim NR(B)] or because he is blameworthy for that belief's leading him to issue the reprimand [so \sim NR(B \supset R)]. That is, if someone denies the conclusion of this Transfer NR argument, it would be natural to assume that that person denies one of the premises – a strong indication of validity. 5

Here is a second example; call it 'Tornado':

Jones sees a tornado approaching. He deliberates for a few seconds as to what to do, and then decides, in order to save his life, to get into his car and drive away, and he proceeds to do so.

Using 'S' to stand for 'Jones sees a tornado approaching', and 'D' to stand for 'Jones drives away to save his life', we can construct the following argument:

$$NR(S)$$
, $NR(S \supset D) / :: NR(D)$

This argument is an instance of Transfer NR, and its premises, as well as its conclusion, seem to be true. Furthermore, Jones qualifies as a normally functioning agent. Again, we think that application of the 'reverse argument' strategy would elicit the intuition that anyone who claimed that the Jones *is*

⁵ Although in the example in the text we construe responsibility in terms of blameworthiness, the example (and the one to follow) would also work for other modes of moral responsibility, such as praiseworthiness. In that case as well, if someone

denies the conclusion, she will have to deny one of the premises, a result that again indicates validity.

responsible for D must believe that Jones is responsible either for S or for $(S \supset D)$; that is, it would elicit the intuition that this instance of Transfer NR is valid.

Thus, we have two examples of intuitively valid Transfer NR arguments that involve causal chains that pass through normally functioning agents. Hence, McKenna's complaint against DA is no longer justified.

4. The dialectical situation

Another way to respond to McKenna's argument, independent of our above response, 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 and DA 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.

We view the dialectic as follows: In an attempt to show that compatibilism is false, van Inwagen introduces a new principle that differs from the principles to which incompatibilists standardly appeal. He first shows that this new principle holds for uncontroversial cases. Then, once the principle is *prima facie* established, he applies it to the controversial cases and draws his anti-compatibilist conclusion. McKenna demands that van Inwagen, *before* he applies his principle to the controversial cases, must first *establish* that his principle holds also in controversial cases. We think that this demand is unreasonable.

4.1 Establishing the prima facie validity of Transfer NR

Let's now move from the abstract to the more concrete and ask how, exactly, van Inwagen proceeds to achieve his goal. First, let's clarify the role played by the examples Snakebite and Plato in his argument. Consider Snakebite, letting 'C' stand for 'John was bitten by a cobra on his thirtieth birthday', and 'D' for 'John died on his thirtieth birthday':

NR(C), $NR(C \supset D) / :: NR(D)$.

The first thing to note is that van Inwagen was not trying to generalize inductively, from this 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 the reverse argument to Snakebite. Suppose Betty claims that someone *is* responsible for John's dying on his thirtieth birthday [so \sim NR(D)]. We ask Betty whether she thinks that someone set the snake on John, and therefore is responsible for the snake's biting John [so \sim NR(C)]. She says 'No'. We then ask her whether she thinks that someone was in a position to stop the snakebite from killing John – e.g., 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 [so \sim NR(C \supset D)]. She again answers 'No'. Betty's claims are 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 Betty'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. For if a rule of inference, or an argument form, is valid, then it is *always* valid. In Transfer NR, van Inwagen has at his disposal a new principle which is distinct from the principles to which

⁶ 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.

incompatibilists standardly appeal. The new principle (which Transfer NR expresses in the form of a rule of inference) is that *non-responsibility transfers across a material conditional*, or, more precisely, that non-responsibility transfers from antecedent to consequent across a material conditional for (the truth of) which no one is responsible.⁷ 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, as van Inwagen claims, on the person who denies the unrestricted validity of Transfer NR.

4.2 Who bears the burden of proof?

McKenna, however, argues that van Inwagen, and incompatibilists generally, still bear the burden of proof. That is, he claims that *before* the incompatibilist can justifiably apply Transfer NR to cases involving normal agency, he must *show* that Transfer NR is valid *in such cases*, and not merely in uncontroversial cases, such as Snakebite and Plato. In other words, McKenna claims that the compatibilist is justified in rejecting the application of Transfer NR to cases involving normal agency, until it is *shown* that Transfer NR holds also in these cases.

In this debate over burden of proof, we side with van Inwagen. We think that once a principle of inference has been established as valid in uncontroversial cases, it can, *prima facie* at least, justifiably be applied to controversial cases as well; that is, the burden of proof is on those who oppose this wider application.

It is instructive to compare van Inwagen's strategy to a similar strategy employed by Peter Singer in another debate. In his well known article 'Famine, Affluence, and Morality' (1972), Singer argued that we ought to donate much more to famine relief than most of us do, or even think we ought to. He begins by establishing the following principle: If it is in our power to prevent something bad

⁷ Note that this is not the same as the principle that non-responsibility is preserved through *causal determination* by factors for which the agent was not responsible.

from happening, without thereby sacrificing anything of comparable moral importance, we ought, morally, to do it. In support of this principle, he cites an uncontroversial example: if I am in a position to save a child who is drowning in a shallow pool, where my only sacrifice would be muddying my clothes (and a small amount of time and effort), then I ought to do it. Having thus elicited intuitive approval of his principle, he then applies the principle to the much more controversial case at hand, arguing that we ought to give our money toward saving lives and alleviating intense and prolonged suffering in far off places, even to the point of 'reducing ourselves to the level of marginal utility.' Whether or not one considers this argument conclusive, there is no question that it has some force. Singer's opponents might argue against him; but the point here is that if they want either to deny or to restrict his principle, they would have to *argue* — which means that Singer has shifted the burden of proof onto his opponents. Of course there are differences between the uncontroversial case (the drowning child) and the controversial case (famine relief); the burden is to explain how the differences are relevant, why they make a difference with respect to the principle in question.

Similarly, once the validity of Transfer NR has been *prima facie* established on the basis of uncontroversial cases, the burden of proof rests on the person who opposes its application to the controversial cases, i.e. the compatibilist. The only two options available to her at this point are, as van Inwagen pointed out, either [1] to propose a plausible alternative account of the intuitive validity of Snakebite and Plato, one that leaves open the possibility that applications of Transfer NR to the controversial cases would be invalid, ⁸ or [2] to provide a non-controversial counterexample to Transfer NR.

4.3 Has McKenna satisfied van Inwagen's dialectical requirement?

0

⁸ 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.

McKenna does not satisfy van Inwagen's [2]. However, one might suggest that implicit in McKenna's discussion is a proposal that satisfies van Inwagen's [1]; 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 2008: 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 cases of normal agency from its range of application. Note, however, that cases of normal agency are precisely the controversial cases, the cases over which the incompatibilist and the compatibilist disagree. Note further that 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 suggests is, in effect, that compatibilists should say that Transfer NR is valid, except where it conflicts with compatibilist intuitions – that is, except where it threatens compatibilism. To us, McKenna's suggestion seems *ad hoc* and dialectically improper.

One might, more sympathetically, interpret McKenna's compatibilist as employing G. E. Moore's well-known strategy of 'inverting the argument'. Moore (1953: 136-38) could not find any flaw in Hume's argument for scepticism about material objects; but he asserted that he is more certain that he does know that there are material objects than he is that Hume's argument is sound, so he will continue to believe that he knows there are material objects, and that somehow Hume's argument must be unsound. Perhaps McKenna's compatibilist can be construed as arguing, similarly, that she is more certain, e.g., that Bush is praiseworthy for getting down and making his daughter happy than she is of the cogency of DA; so she will continue believing that Bush is praiseworthy, and generally that moral responsibility is compatible with determinism, without bothering to point out the flaw in DA.

We have two things to say about this. First, employing Moore's strategy does not constitute a fully satisfactory philosophical response to a philosophical argument; it is just a way of 'learning to live with it.' Presumably Moore would have much preferred to be able to find and exhibit a flaw in Hume's

sceptical argument. As it is, the challenge posed by Hume's argument still stands; and so does the challenge posed to the compatibilist by DA. Secondly, Moore's strategy is most appropriate in cases where virtually everyone finds the conclusion of the argument to be shocking or unacceptable. In Moore's case, we all believe very strongly, at least at an intuitive level, that we know that there are material objects; but the philosophical world is divided over whether determinism is compatible with moral responsibility. Therefore, in using Moore's strategy, McKenna's compatibilist would be seen as stubbornly maintaining her tendentious position in the face of a broadly appealing argument in which she cannot find a flaw. This would be dogmatic and dialectically improper.

5. Conclusion

We have provided two separate responses to McKenna's dialectical objection to Transfer NR. First, our examples Reprimand and Tornado effectively meet his challenge to provide intuitively valid examples of Transfer NR that involve normal agency. Secondly, we have challenged his assessment of the dialectical situation, showing that he seems not to have appreciated the dialectical role of van Inwagen's examples in establishing the *prima facie* validity of Transfer NR.⁹

> Ira M. Schnall Bar-Ilan University Ramat-Gan 52900 uschnall@yahoo.com

David Widerker Bar-Ilan University Ramat-Gan, Israel 52900 widerd@mail.biu.ac.il

⁹ We would like to thank Ori Beck, Carl Ginet, David Gottlieb, Michael Pauen, and Derk Pereboom for helpful comments on earlier versions of this paper.

References

Fischer, J. and M. Ravizza, 1998. Responsibility and Control: A Theory of Moral Responsibility.

Cambridge: Cambridge University Press.

Haji, I. 2008. Reflections on the incompatibilist's Direct Argument. Erkenntniss 68: 1-19.

McKenna, M. 2008. Saying good-bye to the Direct Argument the right way. *Philosophical Review* 117: 349-83.

Moore, G. E. 1953. Some Main Problems of Philosophy. New York: MacMillan.

Rowe, W. L. 1980. Fatalism and truth. Southern Journal of Philosophy 18: 213-19.

Singer, P. 1972. Famine, affluence, and morality. *Philosophy and Public Affairs* 1: 229-43.

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

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