

FREE WILL AND THE BURDEN OF PROOF

William G. Lycan
University of North Carolina

1. Here are some things that are widely believed about free will and determinism.

(1) Free will is *prima facie* incompatible with determinism.

(2) The incompatibility is logical or at least conceptual or a priori.

(3) A compatibilist needs to explain how free will can co-exist with determinism, paradigmatically by offering an analysis of 'free' action that is demonstrably compatible with determinism. (Here is the late Roderick Chisholm, in defense of irreducible or libertarian agent-causation: 'Now if you *can* analyze such statements as "Jones killed his uncle" into event-causation statements, then you may have earned the right to make jokes about the agent as cause. But if you haven't done this, and if all the same you do believe such things as that I raised my arm and that Jones [sic] killed his uncle, and if moreover you still think it's a joke to talk about the agent as cause, then, I'm afraid, the joke is entirely on you.'¹)

(4) Free will is not impugned by quantum indeterminism, at least not in the same decisive way that it is impugned by determinism. To reconcile free will with quantum indeterminism takes work, but the work comes under the heading of metaphysical business-as-usual; to reconcile free will with determinism requires a conceptual breakthrough.

And listen to Laura Waddell Ekstrom on the burden of proof:

...in the absence of an argument to the contrary, it is straightforwardly clear to most all of us as we adopt the practical deliberative point of view toward our own future that the following is true: I am free in what I do at the next moment only if I am not necessitated to do just what I do by the past and the natural laws.... The compatibilist, then, needs a positive argument in favor of the compatibility thesis.²

What is interesting is that each of these claims is tacitly granted by many free-will compatibilists as well as (obviously) by incompatibilists. The compatibilists take up the burden and labor to overcome the supposed conceptual obstacles.

I maintain that (1), (2) and (3) are just false. I believe (4) is false as well--though perhaps not *just* false.

It's actually worse than that. Here is a confession: I have always been free-will-blind. I don't get it. I don't see what the big problem is supposed to be. (My talk today may have the unintended effect of persuading some of you that that's right: that *I just don't get it*.) I am a natural-born, cradle compatibilist. Ted Honderich suggests^{<3>} that even if we are Soft Determinists, each of us still *desires* a measure of libertarian agent-causation, or can easily be made to; the real problem is attitudinal. But if I have such a desire, it operates at a level deeper than introspection can penetrate.

Of course I know that there are serious arguments for incompatibilism. Some are simple, some more complicated, some very complicated and ingenious. But I have never found any of them at all convincing. My purpose in this paper is to try to show that that response is right and proper, and that there are good general dialectical reasons for rejecting all incompatibilist arguments. (To my title I might have added the subtitle, 'in which it is shown that compatibilism is not only true, but the only position rationally available to impartial observers of the issue.')

Ground rules: For the sake of argument, I shall assume that determinism is true. I myself believe that it is not true; but let us assume it, because my main thesis is the compatibility of freedom *with determinism*, and because if I am right in rejecting (4), indeterminism would not help anyway. Also, if no incompatibilist argument succeeds, there is no reason to think that indeterminism would make us any freer.

By 'free' I shall mean free in whatever sense is germane to moral responsibility, that is, such that one is morally responsible for only those of one's actions that were free actions. (I know this usage is not inevitable.^{<4>})

2. I begin with a general methodological point about modality: Compatibilism, not just about free will but generally, on any topic, is the default. For any modal claim to the effect that some statement is a necessary truth, I would say that the burden of proof is on the claim's proponent. A theorist who maintains of something that is not obviously impossible that nonetheless that thing *is* impossible owes us an argument. And since entailment claims are claims of necessity and impossibility, the same applies to them. Anyone who insists that a sentence S1 entails another sentence S2 must defend that thesis if it is controversial. If I tell you that 'Pigs have wings' entails 'It snows every day in Chapel Hill,' you need not scramble to show how there might be a world in which the first was true but the second false; rather, you would rightly demand that I display the alleged modal connection. And of course the same goes for claims of incompatibility.

The point is underscored, I think, if we understand necessity as truth in all possible worlds. The proponent of a necessity, impossibility, entailment or incompatibility claim is saying that *in no possible world whatever* does it occur that so-and-so. That is a universal quantification. Given the richness and incredible variety of the

pluriverse, such a statement cannot be accepted without argument save for the case of basic logical intuitions that virtually everyone shares.

3. Let us turn specifically to (1) and (2). The second thing to notice here is that ‘incompatible’ cannot mean, *logically* incompatible. No one can start with determinism and derive in the predicate calculus any reasonable translation of the denial that anyone’s action is ever free. What ‘incompatible’ must mean is, jointly incompatible with some further principle. I have argued elsewhere<5> that such a further principle is likely to be speculative and suspiciously philosophical, and I shall develop that point below.

The principle cannot be just the truth-functional conditional from determinism to the negation of free will or vice versa, because the incompatibilist is saying more than that determinism and free will do not in fact both obtain. The principle must be independently motivated and entail that they do not both obtain. I shall argue below that any such principle will be dubious.<6>

Not so fast, though. It is not quite so obvious that ‘incompatible’ cannot mean logical incompatibility. In response to my previous work aforementioned, James Tomberlin has made a clever and doughty move.<7> Taking my allusion to the predicate calculus at face value, he pointed out that there are richer logics, notably modal logics, which reveal more logical incompatibilities than can be demonstrated in predicate logic. $[box]P \ \& \ \sim P$, $[diamond]P \ \& \ [box]\sim P$, and $P \ \& \ \sim[diamond]P$ are all logical contradictions within the meaning of the act even though they cannot even be directly formulated in the predicate calculus. (Of course they can be *translated into* the predicate calculus by appeal to an ontology of possible worlds, and in that derivative sense shown valid by use of predicate logic.) Thus, for all I have shown, it is possible that determinism and freedom can be formulated in such a way that they prove to be logically, because modal-logically, incompatible after all, and no extraneous principle, dubious or not, will be needed.

Here, in brief, is how Tomberlin’s argument goes (pp. 128-130). He assumes for *reductio* that ‘Although Bob did A freely, his doing A has a determining cause [C].’ He gradually translates that assumption into the vernacular of possible worlds, and proceeds to derive both ‘[F]or any possible world, if C obtains, Bob does A [in that world]’ and ‘[T]here is some physically possible world such that although C obtains, Bob does not do A,’ which predicate-calculus contradiction reduces the original assumption to absurdity. The derivation proceeds, concessively, through a premise corresponding to the traditional compatibilist hypothetical analysis of ‘could have done otherwise,’ though that premise is not itself translated into possible-worlds talk. So Tomberlin’s thesis is that even if that ostensible pillar of compatibilism is granted, the original assumption is still refuted via the predicate calculus and incompatibilism is thus established.

The trouble with Tomberlin’s argument, I have argued elsewhere,<8> is that

although each of its premises is perfectly true and one might casually think that the conclusion follows from them, the tendentiousness lurks in the untranslated premise, which reads: 'In C Bob could have done otherwise =df In C, if he had chosen to do otherwise and..., then Bob would have done otherwise' (p. 129, ellipsis original). To make the argument demonstrably valid, we have to finish the job and explicate that premise in terms of possible worlds. If we do so in the normal way, according to roughly Stalnaker-Lewis 'similarity' semantics and making the most natural judgments of similarity, the resulting formula does not combine with the previous premises to entail the needed contradiction. (If we suspect that the hypothetical analysans should be treated as a backtracker, <9> and apply a natural similarity analysis for backtrackers, the resulting translation leaves the argument even more obviously invalid.) So Tomberlin has failed to deduce a contradiction from the original assumption.

A similarly a priori and very ingenious modal incompatibilist argument has recently been devised by Ted A. Warfield.<10> There is some question whether the argument is in fact valid even as already formulated into modal notation,<11> but in any case, two English sentence-schemata originally offered by Warfield have to be translated into modal notation in such a way as to make it valid. (The schemata are 'P is true and there's nothing anyone is free to do in the circumstances that even might result in ~P' and 'P is true and there's nothing anyone is free to do in the circumstances that would definitely result in ~P' (p. 173).) I would contend that the translations needed to make the argument valid would have to be tendentious in the characteristically philosophical way, though I cannot argue that here (or even sketch Warfield's complex modal argument).

Thus, the incompatibilist's tendentious assumption need not take the form of a principle (though one can always reconstruct it as such). In these would-be purely modal arguments it can take the form of either a failure to translate in detail, leaving a gap in the argument, or a highly disputable translation.

So, vs. (1) and (2): (1) is false because nothing is 'prima facie incompatible' with anything unless there is a marked air of *logical* contradiction. (2) *might* yet be shown to be true, but induction over spirited and ingenious attempts by excellent philosophers argues otherwise.

4. Against thesis (3): Incompatibilists often challenge compatibilists to say what 'free action' means, if it does not mean an action that is physically undetermined. (Recall our opening quote from Chisholm.) And compatibilists have leaped to respond; Stace and Ayer, for example, offered their infamous hypothetical analyses of 'free action',<12> according to which my action was free iff, roughly, had I wanted / chosen otherwise, I would have done otherwise. Such analyses were promptly attacked by Austin, Chisholm, Keith Lehrer and others.<13>

I think the compatibilists have made a big strategic error. If my previous points,

against (1) and (2), are right, the compatibilist has no obligation to offer any competing *analysis* of ‘free action’. I believe the compatibilists should have balked, and merely insisted that *there is* a perfectly good sense of ‘free’ in which we act freely despite our actions’ being the determined result of pre-existing conditions.

Notice, I said they *should* have balked, not just that they would have been within their dialectical rights to have done so. Notice particularly that in capitulating and consenting to offer analyses, they opened themselves to a kind of attack whose dialectical force is weak but whose rhetorical force is strong: Their opponents could and did vigorously attack the analyses, thereby making it look as though there were something wrong with the compatibilist position in itself.

Think about it: On *any* philosophical topic, the person who propounds an analysis is going to get creamed. Philosophical analyses virtually never work, but are lacerated by counterexample after counterexample. So by agreeing to propose a particular analysis of ‘free action’, our compatibilist is entering a contest s/he cannot hope to win—*not* because there is anything wrong with compatibilism or because ‘free’ really does mean whatever the incompatibilist thinks it does, but only on the entirely general grounds that in the game of philosophical analysis, the analysts’ opponents nearly always win.

Compare J.J.C. Smart and his famous ‘topic-neutral translations’ of mental ascriptions.<14> They were offered in response to each of three objections to Place and Smart’s Identity Theory of mind: the claim that mental ascriptions simply entail the existence of nonphysical states or events, the claim that the Identity Theory violated Leibniz’ law, and the more complicated ‘Objection 3,’ attributed to Max Black in Smart’s footnote 13. Smart contended that mental ascriptions are topic-neutral, in that they entail neither that the states and events ascribed are nonphysical nor that they are physical. (Never mind whether this would in fact have blocked the second and third objections.) Smart sought to show that mental ascriptions are topic-neutral by providing synonyms or paraphrases of them that are both adequate as paraphrases and obviously topic-neutral. (Notoriously, ‘I see a yellowish-orange after-image’ became ‘There is something going on which is like what is going on when I have my eyes open, am awake, and there is an orange illuminated in good light in front of me, that is, when I really see an orange.’)

Allies and critics alike seemed to find this entirely appropriate. David Lewis and David Armstrong offered plainly topic-neutral meaning analyses of their own; Michael Bradley, Frank Jackson and other critics attacked Smart’s paraphrases, to good effect.<15> But, first, why should we expect that an English expression that is in fact topic-neutral must also have a distinct synonym in English that is more obviously topic-neutral? Second, as above, why should we expect philosophical analysis here to produce greater consensus than it practically ever does? And, third, remember our dialectical points against (1) and (2): In a controversial case, non-entailment—hence, here, topic-neutrality—is the default. The burden is on Smart’s

opponent to show that mental ascriptions do entail the existence of nonphysical items.

Smart made the strategic (not philosophical) error of venturing onto extremely dangerous ground when he was quite safe where he was. So too the free-will compatibilist who offers an irenic analysis of 'free action'. Let's not do it. And let's not be bothered by the failures of others' analyses. Chisholm was wrong: There is no joke on us.

5. Against (4): If (1)-(3) are not true, then there is no conceptual problem about free will and determinism; so there is no further asymmetry to underwrite (4). (I nearly wrote, 'then there is no prima facie problem about free will...', but that would have been silly. Of course there is a prima facie problem about free will and determinism, or so much would not have been written about it and we would not be holding this session today. What there isn't is a conceptual problem.)

The reason I earlier expressed doubt about whether (4) is *just* false is the following line of reasoning. Just suppose, contrary to this paper's contention, that free will and determinism are incompatible, so that if determinism is true then there cannot possibly be free will. Then, despite standard claims that in that case interpolating physical randomness would not help, it is still theoretically possible to work out a libertarian theory of agent-causation sitting atop physical indeterminism, that would allow and account for free will. So in principle, indeterminism still goes better with free will than does determinism. (I am not persuaded by that argument, but I do not want to pursue the matter here.)

6. Now I am going to offer a Moorean argument for compatibilism, much the sort of argument Moore would use against idealists and other anti-realists and skeptics.<16> In considering an anti-realist view, he would first derive from it a very specific negative consequence regarding his own everyday experience. For example, take the idealist claim that there are no material objects. From it, Moore would deduce that he himself had no hands--hands being clear cases of material objects.

Of course the idealist had defended the nonexistence of material objects. Let's suppose that the defense had taken the form of a deductively valid argument. The argument must of course have had ultimate premises, themselves undefended. So it is an argument that looks schematically like this:

(P1)

(P2)

.

.

[steps]

.

.

\ (C) There are no material objects. QED

—to which we may add as a corollary,

\ (C') I do not have hands.

We are supposing the argument to be valid. But that is to say only that each of the sets $\{P_1, \dots, P_n, \sim C\}$ and $\{P_1, \dots, P_n, \sim C'\}$ is inconsistent. The idealist of course wants us to accept the premises and therefore to accept C and C' on the strength of them. But nothing about the argument itself forces us to do that, since if we wish to deny its conclusion we have only to reject one of the premises. Any argument can be turned on its head.

Elsewhere^{<17>} I have argued more generally that no deductive ‘proof’ can be anything more than an invitation to compare plausibility: Of the propositions P_1, \dots, P_n , and $\sim C'$, which is the least plausible or credible? The proof affords no deeper investigation.

Applying the crucial question of plausibility comparison to any specific argument for idealism concerning the external world, Moore thought its answer was painfully obvious. Since the reality of material objects is directly entailed by something Moore already knows to be true, that he does have hands, the culprit must be one of the other members of the inconsistent set; it must be one of the premises that is false. It may be interesting to try to decide which one, but that is not necessary in order to vindicate our common-sense belief in the reality of material objects.

To put it that way sounds arbitrary and question-begging, and of course Moore has been widely accused of both faults. Why should he get to choose $\sim C'$ over the argument’s premises and protect it against them? That comes close to merely announcing that the idealist is wrong.

The nonrhetorical answer to the foregoing rhetorical question is this: At least one of the argument’s premises is sure to be distinctively abstract and philosophical, accepted only because it somehow appeals to the idealist. And remember that a deductive argument is only a plausibility comparison. In this example, the comparison is between (a) ‘Here is one hand and here is another’ and (b) a purely philosophical premise such as McTaggart’s assumption that every existing thing has proper parts that are themselves substances. You be the judge. How *could* a proposition like (b) be considered as plausible as (a)? How could you possibly be more confident that every existent thing has proper parts that are substances, than that you have hands?

The epistemic credentials of metaphysical premises (often called ‘intuitions’) are obscure. It is hard to say why a given metaphysician should be strongly attracted to a particular such premise, such as (b), when doubtless we can find another metaphysician--perhaps only in another part of the world or another era, but possibly

just next door--who firmly rejects it. By contrast, Moore has excellent grounds for the competing proposition (a): He remembers seeing and feeling his hands on millions of occasions, and he can do so again at will. A forced choice between (a) and (b) has got to favor (a).

7. Before we get back to free will, here are six important caveats regarding Moore's technique: First, whatever one now thinks of Moore's response to the idealist—in particular, even if one is not convinced of the plausibility comparison—Moore has not *begged the question* against the idealist. In judging that one of a pair of propositions is more credible than the other, one may be mistaken, but it is no dialectical offense (unless, perhaps, the other proposition is just the negation of the first, which is not the case here).

Second, it would not help the idealist to claim that her/his philosophical premise is *analytic*, or a somehow “conceptual” truth, and so not in need of defense. Even if one is unpersuaded by Quine of the nonexistence of such truths, the appeal will not help the idealist here, for no one who reasonably thinks that the premise is false is going to be converted by the bare assertion that the principle is conceptually true.

Third, contra many of Moore's critics over the years, Moore is not clinging to his commonsense beliefs come what may and treating common sense as sacred and invulnerable to criticism. Moore never held that common sense is irrefutable. Common-sense beliefs can be corrected by careful empirical investigation and scientific theorizing, which is what happened in, e.g., the case of the earth's shape and motion. The point against the idealist is that *philosophers* are not empirical investigators or scientists. McTaggart provided no evidence for his claim that every existent thing has proper parts that are substances; it just seemed true to him, for some reason. Though common sense must yield to evidence, it need not yield to bare metaphysical pronouncement.<18>

Fourth caveat: It's important to see that the argument contains no premise *about* commonsense propositions themselves, e.g., not that we should give commonsense beliefs considerable weight, or that commonsense beliefs are *prima facie* justified, or that (good God) they have a ‘right of ancient possession’<19>. It is not that ‘I have hands’ etc. are known or justified in virtue of their being commonsense propositions. It is just that they are individually more plausible than are the premises of any philosophical argument intended to show that they are false. (A later philosophical *explanation* of their plausibility might advert to their being commonsensical, though I myself take a different line.)

Fifth, do not be too quick to dismiss Moore's method as shallow and superficial, disrespectful of our need for a deep philosophical critique of common sense. There is an old and well-entrenched idea that philosophy can get above, or beneath, the body of belief constituted by common-sense-plus-science, and subject that body as a whole to deeper rational examination. I reject that idea, essentially on the grounds of what I

have said earlier about the unprobativeness of deductive arguments; I have criticized it more extensively elsewhere.<20>

And, sixth: To make my point, I do not need a *theory of plausibility/credibility*. Granted, the psychological basis and normative authority of ‘plausibility’ judgments are important philosophical issues, but they are just that: philosophical issues. Actual plausibility comparisons made in real life do not depend on having a well justified theory of plausibility—else only professional epistemologists could make them. That people often ride the bus is more plausible or credible than that a BMW will do over 120 mph, or that more people drive vintage Jaguars than drive Chevrolets, or that there is life somewhere else in our galaxy, or that the cause of an idea must have at least as much formal reality as the idea itself has objective reality. I do not have to have any philosophical theory of plausibility in order to be entirely justified in making such judgments, any more than I have to have a philosophical theory of meaning in order to know what my wife has just said.<21>

8. And finally we can state the argument for free-will compatibilism.

I used to claim that a qualified, macro-event-level version of determinism is itself common sense.<22> That has proved to be problematic, for it may be a case in which a commonsense proposition, even the qualified version of determinism, is refuted by compelling philosophical argument based on indeterminist science. But I need not stand by that claim. For purposes of my Moorean argument, all determinism needs to be is: not an affront to common sense.

In any case, the incompatibilist must assume determinism for the sake of argument. And whether or not it is true, numerous commonsense claims of free action always will be more plausible than are the *purely philosophical* premises (or translation lore etc.) of any argument designed to convince us of incompatibilism. So we reject at least one of those premises in each case. Since the incompatibilist argument fails and compatibilism is the default (cf. again the general methodological point, section 2 above), compatibilism rules. Thus, to say the least, it is hard to imagine how compatibilism could be rationally discredited.

9. Two more caveats are immediately needed.

I am not saying that *compatibilism* is a commonsense view. Ekstrom considers such a claim;<23> I am not sure whether she rejects it, though she certainly denies that it is a good reason for embracing compatibilism. I firmly reject it. Compatibilism is a controversial philosophical thesis. (Nor do I buy the ‘*Mind* argument’ that incompatibilist accounts are incoherent.<24>)

Further, in insisting that some of our actions are free, I am not committing the magician’s-assistant fallacy, i.e., taking the fact of failing to introspect determining causes of my actions as a successful introspecting that the actions are free. Nor am I making any other (direct) appeal to phenomenology.

10. I contend, then, that every incompatibilist argument is bound to fail. That is not to say that each will fail in the same way. But to get a feel for the failure, let us take a quick look at one example, an influential and apparently powerful incompatibilist argument, the ‘Consequence argument’ due variously to Carl Ginet, David Wiggins, Peter van Inwagen and James Lamb.<25> For convenience, let us focus on van Inwagen’s 1983 version:<26>

$$\begin{array}{l} [\text{box}]S \\ [\text{box}](S \rightarrow A) \\ \hline \neg [\text{box}]A, \end{array}$$

where the box is interpreted as ‘unalterability’ of some sort, A is the performance of an arbitrarily selected future action of mine, and S is a total efficient cause of A existing before I was born. (Here as always we assume determinism, in this case for conditional proof.) The argument’s appearance of validity is unmistakable. And if we construe the box as, unalterability *by me* in particular, its conclusion denies me free will.

Michael Slote argues that although this inference is attractive because the box is felt to be ‘agglomerative’ ($[\text{box}]X, [\text{box}]Y \vdash [\text{box}](X \& Y)$), agglomerativity fails for some real-life modalities.<27> (He focuses on agglomerativity because he thinks van Inwagen’s inference proceeds by agglomeration followed by closure under necessity, but as we shall see, this interpretation is inessential to the objection.) Consider, e.g., *nonaccidentalness* in the everyday sense. Slote explicates the nonaccidentalness of an event in terms of the event’s being the outcome of what he calls a ‘routine plan’ (RP). (In virtue of the plan, the event was quite normal and to be expected.) Now (p. 15), one day in a bank there is an accidental meeting between Jules and Jim. Each of the two has been sent to the bank ‘as part of a well-known routine or plan of [respective] office functioning.’ Thus, it is no accident that Jules arrives at the bank when he does, and it is no accident that Jim arrives at the bank when he, Jim, does. But it is entirely accidental that both are there at the same time. Jules’ presence in the bank was necessitated by RP1 and Jim’s presence there was necessitated by RP2, but there is no RP that necessitated the simultaneous Jules’-presence-and-Jim’s-presence, because in particular the amalgam RP1+RP2 is not itself a RP.

$$\begin{array}{l} [\text{box}_{\text{RP1}}](\text{Jules arrives at the bank at } t_b) \\ [\text{box}_{\text{RP2}}](\text{Jim arrives at the bank at } t_b) \\ \hline \end{array}$$

--but nothing of the form $[\text{box}_{\text{RPn}}](\text{Jules arrives at the bank at } t_b \& \text{Jim arrives at the bank at } t_b)$ follows, because there is no such RPn. Agglomeration fails.

The case of the Consequence argument’s premises is parallel, except for being a

modalized Modus Ponens rather than modalized Conjunction Introduction. The analogue of an RP is what I shall call a ‘me-excluding determinant.’ Why is S, the total efficient cause of A existing before I was born, unalterable by me? Because it was in place quite independently of my desires, wishes, intentions, for that matter beliefs, and whatever other conative structure would be relevant (hereafter just ‘my CS’), because I neglected to exist at the time. Why is the material conditional $S \rightarrow A$ unalterable by me? Because it is a logical consequence of the laws of nature, and both logic and the laws hold quite independently of my CS. It is in that sense that S and $S \rightarrow A$ are me-excluding determinants (MEDs).

But the individual MEDs differ as between van Inwagen’s first premise, [box]S, and the second, [box]($S \rightarrow A$). We have

$$\frac{[\text{box}_{\text{MED}1}]S}{[\text{box}_{\text{MED}2}](S \rightarrow A)} .$$

But nothing of the form [box_{MED_n}]A follows, because there is no such MED_n. In particular, the amalgam $S \& (S \rightarrow A)$ is not a MED, for it does not determine A in a way that bypasses my CS. Since A is a future action of mine, A is hardly unaffected by my CS. In fact, my CS is an, if not the, prominent element of A’s total efficient cause. To be unalterable by me is to be determined or necessitated in a certain way, viz., in a me-excluding way. S and $S \rightarrow A$ are necessitated in that way, but A, though necessitated in other ways, is not necessitated in that one. The inference-pattern fails.

To put it back in terms of agglomerativity for Slote’s sake: In general, even where Z is entailed by the conjunction $X \& Y$, that X is determined by a MED and Y is determined by a MED does not entail that Z is determined by a MED. So [box]A cannot be derived.

11. ‘Oh, spoken like a good Soft Determinist!’ van Inwagen might reply.<28> Indeed, my moves in this paper against Tomberlin’s and van Inwagen’s arguments may have sounded at every turn as though I had said, ‘Let’s all remember that Soft Determinism is true, and just stiffarm whatever contrary claim we have to in order to save our view.’ Van Inwagen’s argument schema seems valid to him on what he considers any reasonable interpretation of the box; modalities such as mine that fail to support the schema are conspicuously already congenial to the Soft Determinist. Van Inwagen admits that his argument will never convince the die-hard compatibilist,<29> but for his part he cannot hear any Soft-Determinist-leaning interpretation of the box as a reasonable interpretation. It is at best a stalemate.

I believe it is no stalemate, and that Slote and I win. First, our objection to the Consequence argument does not presuppose compatibilism, let alone the truth of Soft Determinism. Our objection is that the argument employs modalities that are, like

virtually every other modality expressible in English, relative modalities, and that on one entirely reasonable interpretation the argument's premises are true and its conclusion false. Granted, the interpretation is guided by a conception of action that is indeed congenial to Soft Determinism. (And I admit here and now that that will be my offstage strategy in responding to every incompatibilist argument.) But it does not follow that the objection presupposes the compatibility of free will with determinism; it does not. Of course, so far as I can see, van Inwagen's controversial inference does not presuppose incompatibilism either, so this first point does not show that we are not in stalemate. But if we are in stalemate, it does not take the form of mutual question-begging.

Second, I have pointed out that, so far as has been shown, the Consequence argument is invalid. To respond, van Inwagen or another incompatibilist would have to either add another premise, or block my counter-interpretation in some principled way. And, as before, the premise or principle thus introduced would have to be more plausible than the commonsensical claim that I did A freely in the sense germane to responsibility. My doubt that the incompatibilist can do that is my reason for denying that we are in stalemate. Remember that the compatibilist bears no corresponding burden; if some conclusion does not obviously follow from some premise, we can only wait to see why the proponents of entailment think it does follow.

In 'Reply to Christopher Hill' (ibid.), van Inwagen offers a sensible dialectical model (p. 58): Think of the issue as a debate conducted by a compatibilist on one side and an incompatibilist on the other, but before an audience that is agnostic on the issue. Each of the two debaters is trying to convert, not necessarily the opposing debater, but the hitherto neutral listeners. So, as the incompatibilist debater, van Inwagen need not restrict himself to premises that would be acceptable to the compatibilist, but must only present an overall case, including replies to the compatibilist's opening points, further rejoinders, etc.) that will or should sway the agnostics. In like wise, the compatibilist may use premises that would be unacceptable to the incompatibilist, so long as they are likely to be granted by the neutral audience.

That is a good test. And the ultimate point of my Moorean argument is to show that the audience will always be forced to choose between, on the one hand, a host of commonsense propositions about their and others' doing things freely, and on the other, a purely philosophical principle that is controversial even among philosophers. In the nature of things, I have argued, philosophy will always lose that one. Compatibilism still rules.<30>

Notes

1 'Comments and Replies,' *Philosophia* 7, Nos. 3-4 (July 1978), 597-636, p. 623.

2 *Free Will* (Boulder, CO:Westview Press, 2000), p. 57.

3 *How Free Are You?* (Oxford: Oxford University Press, 1993); ['Determinism as True, Compatibilism and Incompatibilism as Both False, and the Real Problem'](#).

4 For example, Ekstrom (op. cit, p. 8) eschews it, arguing plausibly that the exact relation between freedom and moral responsibility is unclear and disputable. But the alternative seems to be either metaphor or characterization vapid enough to be all too obviously available to the compatibilist ('our actions are truly attributable to our selves..., ultimately *up to us*' (Ekstrom, p. 3)).

5 [Consciousness](#) (Cambridge: Bradford Books / MIT Press, 1987), Chapter 9.

6 N.b., I will not contend, as some compatibilists have in attacking incompatibilist arguments, that the principle begs the question.

7 'Whither Compatibilism? A Query for Lycan', *Philosophical Papers* 17, No. 2 (August 1988), 127-31.

8 'Compatibilism Now and Forever: A Reply to Tomberlin', *Philosophical Papers* 17, No. 2 (August 1988), 133-139.

9 In the sense of David Lewis' 'Counterfactual Dependence and Time's Arrow', *Noûs* 13, No. 4 (November 1979), 455-76. The idea derives originally from P.B. Downing, 'Subjunctive Conditionals, Time Order, and Causation', *Proceedings of the Aristotelian Society* 59 (1959), 125-40.

10 'Causal Determinism and Human Freedom Are Incompatible: A New Argument for Incompatibilism', in *Philosophical Perspectives 14: Action and Freedom*, J. Tomberlin (ed.) (Atascadero, CA: Ridgeview Publishing, 2000), 167-80, pp.172-77.

11 Michael Kremer, 'How Not to Argue for Incompatibilism', MS, University of Notre Dame.

12 W.T. Stace, *Religion and the Modern Mind* (New York: Lippincott / Harper and Row, 1952); A.J. Ayer, 'Freedom and Necessity', in *Philosophical Essays* (London: Macmillan, 1954), 271-84.

13 J.L. Austin, 'Ifs and Cans', *Proceedings of the British Academy* 42 (1956), 109-32; reprinted in *Philosophical Papers* (Oxford: Oxford University Press, 1961); R.M. Chisholm, 'J.L. Austin's Philosophical Papers', *Mind* 73, No. 289 (January 1964), 1-

26; K. Lehrer, 'Preferences, Conditionals and Freedom', in *Time and Cause*, P. van Inwagen (ed.) (Dordrecht: D. Reidel, 1980), 187-201.

14 'Sensations and Brain Processes', *Philosophical Review* 68, No. 2 (April 1959), 141-156. For discussion of the 'topic-neutrality problem' generally, see Chapter 2 of my [Consciousness](#), loc. cit.

15 D.K. Lewis, 'An Argument for the Identity Theory', *Journal of Philosophy* 63, No. 1 (January 1966), 17-25; D.M. Armstrong, *A Materialist Theory of the Mind* (London: Routledge and Kegan Paul, 1968); M.C. Bradley, 'Sensations, Brain-Processes, and Colours', *Australasian Journal of Philosophy* 41, No. 4 (December 1963); F. Jackson, *Perception* (Cambridge: Cambridge University Press, 1977).

16 The strategy is more fully expounded and defended in my '[Moore Against the New Sceptics](#)', *Philosophical Studies* 103, No. 1 (March 2001), 35-53.

17 *Judgement and Justification* (Cambridge: Cambridge University Press, 1988), Chapter 6.

18 In saying this, I am not assuming a clear distinction between theoretical science and metaphysics, nor do I believe in any such distinction. But that is to say only that there are borderline cases.

19 Keith Lehrer, 'Why Not Skepticism?', *Philosophical Forum* 2, No. 3 (Spring 1971), 283-98 (quoting Thomas Reid).

20 '[Moore Against the New Sceptics](#)', loc. cit.

21 Which is not to say that I don't have such a theory; see *Judgement and Justification*, loc. cit. Chapter 7.

22 [Consciousness](#), loc. cit., pp. 113-14.

23 Op. cit., pp. 56-57.

24 R.E. Hobart, 'Free Will as Involving Determinism and Inconceivable Without It', *Mind* 43, No. 169 (January 1934), 1-27; P.H. Nowell-Smith, 'Free Will and Moral Responsibility', *Mind* 57, No. 225 (January 1948), 45-61; and ; J.J.C. Smart, 'Free-Will, Praise and Blame', *Mind* 70, No. 279 (July 1961), 291-306; also, Ayer, op. cit. I believe the term 'the *Mind* argument' was coined by Peter van Inwagen, in *An Essay on Free Will* (Oxford: Clarendon Press, 1983); he distinguishes three different 'forms' or 'strands' of it.

25 C. Ginet, 'Might We Have No Choice?' in *Freedom and Determinism*, K. Lehrer (ed.) (New York: Random House, 1966), 87-104; D. Wiggins, 'Towards a Reasonable Libertarianism', in *Essays on Freedom of Action*, T. Honderich (ed.) (London: Routledge and Kegan Paul, 1973), 31-61; P. van Inwagen, 'The Incompatibility of Free Will and Determinism', *Philosophical Studies* 27, No. 3 (March 1975), 185-199, and *An Essay on Free Will*, loc. cit.; J. Lamb, 'On a Proof of Incompatibilism', *Philosophical Review* 86, No. 1 (January 1977), 20-35.

26 There are now a number of interestingly different versions of the Consequence argument, subject to somewhat different sets of objections. These are nicely catalogued and discussed by Ekstrom, op. cit., Chapter 2. I criticize van Inwagen's version more extensively in Chapter 8 of *Modality and Meaning* (Dordrecht: Kluwer Academic Publishing, 1994), though that discussion is marred by some vicious copy-editing errors.

27 'Selective Necessity and the Free-Will Problem', *Journal of Philosophy* 79, No. 1 (January 1982), 5-24.

28 Van Inwagen did reply to Slote essentially in that way, in an unpublished note, 'Modal Inference and the Free-Will Problem'.

29 'Reply to Christopher Hill', *Analysis* 52, No. 2 (April 1992), 56-61, p. 58.

30 I thank Jim Tomberlin for his stimulating article cited above. Thanks also to Fritz Warfield for the talk and subsequent conversation that directly inspired this paper. I am grateful to Anthony O'Hear and Tim Crane for putting together 'Free Will day' at University College London, and to the Royal Institute audience for spirited and helpful discussion.