Able to Do the Impossible

Jack Spencer
Princeton University

January 2013
(Please do not quote or cite without permission)

1 Introduction

There is principle that is often taken for granted that connects abilities to metaphysical possibilities. I call it the *poss-ability principle*: If S is able to ϕ, then it is metaphysically possible for S to ϕ. The poss-ability principle states each of an agent's abilities must be 'modally witnessed' by a metaphysical possibility in which the agent exercises the ability. In other words, if S is able to ϕ, then there must be some possible world in which S (exercises her ability and) ϕs.[1]

Much hangs on the poss-ability principle. The leading arguments for incompatibilism have the poss-ability principle as a premise. The leading theories of ability have the poss-ability principle as a consequence. And the poss-ability principle is often a critical assumption operating in the background. The thesis that an object is visible (or knowable) only if there is some possible world in which the object is seen (or known) relies, implicitly, on the poss-ability principle, as does the widely held thesis that 'Ought' implies 'It is metaphysically possible that'. Most philosophers accept the poss-ability principle, of course, which puts me in the minority.

[1] Possible worlds are maximally complete metaphysical possibilities. I assume that incomplete metaphysical possibilities are always contained within complete metaphysical possibilities. Thus, if there is some incomplete metaphysical possibility in which S ϕs, then there is also be some complete metaphysical possibility in which S ϕs.

I reject the poss-ability principle. In my view, it is not a constraint on an agent having an ability that there be some possible world in which the agent exercises the ability. An agent might be able to do what it is metaphysically impossible for her to do. In fact, and this is more surprising, an agent might be able to do what it is metaphysically impossible (not just for her, but *tout court*) to do. The poss-ability principle admits of counterexamples.

I am not the first philosopher to allege that the poss-ability principle admits of counterexamples. The familiar, although much disputed, putative counterexamples to the poss-ability principle are various instances of the Grandfather Paradox (e.g. cases of attempted autoinfanticide), and involve an agent traveling backwards in time.[2] I, however, want to put time travel to one side. It is interesting to speculate about whether time travel is possible, and, if so, what time travelers might be able to do. But whether the poss-ability principle is true does not depend on any deep questions about the nature of time. We should reject the poss-ability principle, I think, even if time travel is shown to be metaphysically impossible. For contrary to what philosophers often suppose, there are counterexamples to the poss-ability principle that do not involve backwards time travel. Such cases have not to my knowledge been discussed hitherto. I will call them *Nate Cases*. The main purpose of this essay is to introduce Nate Cases and explore their ramifications, which I think are far-reaching.

A Nate Case, in a nutshell, is a case in which one possible world (or a class of worlds, even an infinite class) stands apart from the rest of the possible worlds vis-à-vis the prospects of ϕing. If ϕing occurs anywhere in modal space, it occurs exclusively in the one possible world. Agents in the

[2] A much discussed question: if time travel is possible, then might a time traveler, given the right opportunity, be able to kill her younger self, notwithstanding the fact that it is metaphysically impossible for her to kill her younger self. Philosophers disagree. David Lewis thinks so; he thinks that time travel is possible and that relations of diachronic identity do not impose further constraints on what agents are able to do. See his "The Paradoxes of Time Travel," in his *Philosophical Papers, Vol. 2* (Oxford: Oxford University Press 1986) pp. 67-80, 74-80. Kadri Vihvelin thinks not; she thinks that relations of diachronic identity do indeed impose further constraints on what agents are able to do. See her "What Time Travelers Cannot Do," *Philosophical Studies* 91 (1996) 315-330. I am on the fence. See footnote 12.

one possible world might be able to ϕ, but agents in any other possible world are clearly and demonstrably unable to ϕ. As there are unexercised abilities, the question arises whether an agent could have the unexercised ability to ϕ. And I think so. Thus, the blueprint for a Nate Case: the agent—Nate—has the unexercised ability to ϕ in the one possible world, and lacks the ability to ϕ in every other possible world; there is one possible world in which Nate is able to ϕ, but there are no possible worlds in which Nate (exercises his ability and) ϕs.

The most straightforward Nate Cases take place in deterministic universes and rely on the assumption of compatibilism, i.e. that an agent might be able to do what she is determined not to do. Once we get into the swing of things, however, we see immediately that there are indeterministic Nate Cases too.

Are there incompatibilistic Nate Cases? I will leave that an open question. I believe that (most) compatibilists should reject the poss-ability principle, which is a surprising thesis in its own right. It is a harder question whether (most) incompatibilists should reject the poss-ability principle. Tentatively, I suspect so. While incompatibilists will find the majority of the Nate Cases in this essay unconvincing, for the obvious reasons, there is one Nate Case in §4 and two in §6 that I think incompatibilists should find persuasive. Of course, being a compatibilist myself, I think that incompatibilism is false, so I would reject the poss-ability principle even if it turned out that *by the lights of incompatibilism* the poss-ability principle is true. My thesis is that Nate Cases are counterexamples to the poss-ability principle. My task is to convince my fellow compatibilists and the people on the fence. If I can convince some incompatibilists along the way, then so much the better.

2 The Counterexample

Suppose that determinism is true. Let h be the complete specification of the initial conditions of the universe. Let l be the complete specification of the deterministic laws of nature. Let $h\&l$ be their conjunction.

Suppose that Nate, our protagonist, has not, does not, and will not believe that $h\&l$. Nate never finds himself reading a book or listening to a radio program about the initial conditions or the laws of nature; Nate was home from school and sick with the flu on the day that his physics teacher covered the initial conditions and the laws of nature in class, and the physics teacher never bothered to go over the material again; etc. We may suppose that it is fairly common knowledge in Nate's community that $h\&l$, that college-bound high school seniors are expected to know that $h\&l$, that many of Nate's classmates know that $h\&l$, and that Nate is one of the brightest students in his class. The proposition *that* $h\&l$ does not exceed Nate's cognitive wherewithal, either in length or in complexity, and there is no special obstacle preventing Nate from forming the belief. The fact that Nate does not come to believe that $h\&l$ is, as we might put it in ordinary English, an historical accident.

Compatibilists, one and all, will agree that Nate is able to believe the true proposition *that* $h\&l$. I make a stronger claim. I claim that Nate is able to truly believe—or better still—that Nate is able to know that $h\&l$. Nate has a factive ability.

A factive ability is an ability to perform a factive action, such as *knowing that* p, *learning that* q, or *seeing that* r.[3] Nate has many factive abilities, of course. Nate is able to know, learn, correctly predict that, truly assert, regret, be surprised that, realize, discover, see, and remember a great variety of things. My claim is that $h\&l$ is one such. Nate is able to know that $h\&l$; Nate is able to learn, realize, and discover that $h\&l$; if Nate once knew that $h\&l$ but has since forgotten, then Nate is able to remember that $h\&l$, too.

It is, however, metaphysically impossible for Nate to know that $h\&l$, as every possible world in which Nate believes that $h\&l$ is a world in which $h\&l$ is false; and knowledge requires true belief. The poss-ability principle would have us deny that Nate is able to know that $h\&l$; I deny the poss-ability principle, instead. I say Nate is able to know that $h\&l$, no less so than his fellow classmates. (If Nate were not so able, then he should not have been

[3]Perhaps knowing is not an action but a state. In that case, where I say 'know' read 'come to know'.

enrolled in the physics class to begin with!) The difference between Nate and his classmates is a difference not of ability but of opportunity: Nate was absent from the lecture (i.e. the knowledge gaining opportunity) that his classmates attended. But we should distinguish a lack of opportunity, even a necessary lack of opportunity, from a lack of ability. I say that Nate is a counterexample to the poss-ability principle. Nate is able to do what it is metaphysically impossible for Nate to do; namely, know that $h\&l$.

3 The Diagnosis

Nate Cases are more convincing when we gain an appreciation for what makes them tick. To that end, I broach the epistemology of ability.

Abilities are not easy to detect. In general, it is much easier to detect the actions that spring from an agent's abilities than the agent's abilities themselves. For that reason, in the epistemology of ability, we use inference to the best explanation, inferring from facts about what agents *do* to conclusions about what agents are *able to do*. The best evidence that an agent is able to ϕ is the agent repeatedly attempting to ϕ and repeatedly succeeding—Michael Jordan's attempts to slam-dunk a basketball are repeatedly successful, so, by inference to the best explanation, Michael Jordan is able to slam-dunk a basketball. The best evidence that an agent is unable to ϕ is the agent repeatedly attempting to ϕ and repeatedly failing—my attempts to slam-dunk a basketball are repeatedly unsuccessful, so, by inference to the best explanation, I am unable to slam-dunk a basketball.[4]

Keep in mind the order of explanation. It would be wrong to think that

[4] A point of terminology: I use the word 'attempt' in a broad sense to mean: circumstance in which the would-be ability might reveal itself in action. We do not attempt (in the ordinary sense) to sneeze, for example, but we do attempt (in my sense) to sneeze. Having a cold is an attempt (in my sense) to sneeze. An attempt to sneeze is simply a circumstance in which the would-be ability to sneeze might reveal itself in a sneezing. An attempt need not be effortful. The same goes for believing or knowing. We do not attempt (in the ordinary sense) to know, but we do attempt (in my sense) to know. A circumstance in which a trusted source tells the agent that p is an attempt by the agent to believe that p.

the verdicts of our attempts are explanatorily prior to our abilities (as if first someone successfully tries to slam-dunk a basketball and only thereafter, and therefore, is able to jump so high!). Successful or unsuccessful attempts are not what *make* agents able or unable. Rather, it is exactly the other way around: our abilities explain the verdicts of our attempts. Michael Jordan's attempts to slam-dunk a basketball are successes, and mine are not, *because* Michael Jordan is able to slam-dunk a basketball, and I am not.

But abilities are hard to detect, at least in any direct way. You cannot just look at someone and thereby ascertain what he is and is not able to do. So the epistemology of ability runs in the opposite direction of the metaphysics. The metaphysics follows the order of explanation: the verdicts of an agent's attempts are explained by what an agent is and is not able to do. The epistemology follows the order of indication: the verdicts of an agent's attempts indicate what an agent is and is not able to do. Successful attempts indicate ability, and failed attempts indicate inability.

The method of determining what agents are able to do by looking to the verdicts of their attempts would suffer from a poverty of stimulus if we restricted our attention to present or actual attempts, so we take into consideration non-present and non-actual attempts as well. To keep track of things, we index abilities to times and worlds. After all, what an agent is able to do changes over time and across possibilities. The question that we want to answer is whether the agent is presently and actually able to ϕ. Our strategy for answering the question is to look to the agent's attempts to ϕ—both present and non-present attempts, both actual and non-actual attempts—and extrapolate from the verdicts of those attempts to our conclusion.

Of course, not all of an agent's attempts are on an evidential par. Some attempts are epistemically relevant, but others are irrelevant. To dramatize the point, suppose that in the future I sprout wings. Making the most of my newfound pinions, I repeatedly attempt to fly and repeatedly succeed. Those are indeed repeated and successful attempts by me to fly. But everyone recognizes that my future (winged) attempts to fly do not provide good evidence about what I (being unwinged) am presently and actually able to

do. That my (winged) attempts to fly are successes says little or nothing, or something misleading, about whether I am presently and actually able to fly. My (winged) attempts are not, as I will put it, *representative* attempts.

Which are the representative attempts? Hold fixed the facts that together determine whether S is presently and actually able to ϕ, and consider the attempts by S to ϕ at times and at worlds at which those facts obtain—*those* are the representative attempts. The attempts by S to ϕ that are representative of whether S is presently and actually able to ϕ are the attempts done by versions of S that are neither enhanced nor diminished (as compared to the present and actual version of S) in circumstances that are neither more nor less favorable to S ϕing (as compared to the present and actual circumstances).[5]

The epistemology of ability goes via the verdicts of the representative attempts. The verdicts of the unrepresentative attempts carry little or no information; we can ignore them. It is the verdicts of the representative attempts that reveal what an agent is and is not able to do. Successful representative attempts indicate ability, and failed representative attempts indicate inability.

There is a point here to make about the relationship between abilities and counterfactuals. I submit that the relationship is primarily epistemological, that counterfactual streamline the epistemology of ability by providing a quick and for the most part reliable way of identifying representative attempts.

We want to know whether the agent is presently and actually able to ϕ. So consider the closest possible worlds in which the agent attempts to ϕ. What are the verdicts of the agent's *closest* attempts? Are they successes or failures? For the most part, if an agent's closest attempts are successes,

[5] I offer a binary distinction between representative attempts and unrepresentative attempts, but this is somewhat artificial. There is a gradation: one attempt may be more or less representative than another. My distinction may be understood to separate the perfectly representative attempts from the rest. There are in-between cases. For example, a successful attempt by S to ϕ done by a slightly diminished version of S, or in slightly less favorable circumstances, is somewhere in-between being a perfectly representative attempt and being an unrepresentative attempt.

then the agent is able, and if the agent's closest attempts are failures, then the agent is unable. And it is easy to explain why the counterfactual test delivers what are for the most part the right results: in most cases, an agent's closest attempts are also *representative* attempts. That is, in most cases, an agent's closest attempts are done by versions of the agent that are neither enhanced nor diminished in circumstances that are neither more nor less favorable. Close attempts thus function as easy-to-identify proxies for representative attempts.

The counterfactual test has its well-known limitations, however. When an agent's closest attempts are representative attempts, then the counterfactual test works fine. But when an agent's closest attempts are unrepresentative attempts, the counterfactual test falters. There are two sorts of problems.

(1). *The verdicts of the closest attempts might differ from the verdicts of the representative attempts.* Consider a finkish agent, e.g. me. I have puny biceps, and I am hopelessly unable to do ten chin-ups. But suppose that I have a fairy angel who hates to see me fail. If I tried to do ten chin-ups, then my fairy angel would intervene, temporarily beef-up my biceps, and my attempt would therefore be a success. Let us stipulate that the closest worlds in which I attempt to do ten chin-ups are worlds in which my fairy angel intervenes and supplements my strength. Good for me, I suppose, for counterfactually doing so many chin-ups! But (beefed-up) attempts by me to do ten chin-ups are clearly unrepresentative of what I (being puny) am presently and actually able to do. It may be true that if I tried to do ten chin-ups, I would succeed. It may be true that if I tried to do ten chin-ups, I would magically and temporarily gain the ability to do ten chin-ups. But despite all of that, I am not presently and actually able to do ten chin-ups. My closest attempts are successes, but my closest attempts are not representative attempts.

There are representative attempts by me to do ten chin-ups, of course. They are attempts done by *unenhanced* versions of me, i.e. versions of me that have my same puny biceps, and they are, for the most part, failed attempts. (As I said, I am unable to do ten chin-ups.)

If there are representative attempts, then we can mitigate and even eliminate the difference between the verdicts of the closest attempts and the verdicts of the representative attempts by supplementing the antecedent. We began with the first counterfactual test: if S tried to ϕ, then S would or might succeed. We do better by using the second counterfactual test: if S remained intrinsically very much the same and tried to ϕ, then S would or might succeed. The second counterfactual test is better than the first because, while it is frequently the case that the closest attempts by S to ϕ are representative attempts, it is even more frequently the case that the closest attempts by a not-too-intrinsically-dissimilar version of S to ϕ are representative attempts. For example, if the closest worlds in which I try to do ten chin-ups are worlds in which my fairy angel supplements my strength, then the first counterfactual test delivers the wrong results, but the second counterfactual test might still get it right.

But the second counterfactual test is still imperfect, and for the same basic reason: the closest attempts by a not-too-intrinsically-dissimilar version of the agent are not necessarily representative of what the agent is presently and actually able to do. Imagine that my fairy angel operates not by changing my intrinsic nature but my altering my circumstances. When I try to do ten chin-ups, my fairy angel leaves my puny biceps alone but produces a propitious gust of wind. My (wind-assisted) attempts are no more representative than are my (beefed-up) attempts.

We can continue in this fashion, supplementing the antecedent bit by bit, doing an ever better job of ensuring that the closest, antecedent-satisfying attempts are representative attempts, and thus improving the counterfactual test. For example, we can do better still by using the third counterfactual test: if S remained intrinsically very much the same and the circumstances remained very much the same and S tried to ϕ, then S would or might succeed. If we want to cut to the chase, then we should simply stipulate that the closest attempts are representative attempts.[6] Consider the fourth

[6] If we want a reductive analysis of ability, then it is may be illegitimate to appeal to representativeness, for representativeness might itself be properly understood in terms of ability. But I am not offering an analysis, let alone a reductive one, so my appeal to

counterfactual test: if S made a representative attempt to ϕ, then S would or might succeed.[7] The fourth counterfactual test ensures that if there are representative attempts, the closest attempts are among them. This is a dramatic improvement; it eliminates any potential difference between the verdicts of the closest attempts and the verdicts of the representative attempts. In my opinion, the fourth counterfactual test of ability is, qua test of ability, as good as a counterfactual test can be.

But even the fourth counterfactual test is imperfect. The problem is not that the verdicts of the closest attempts might differ from the verdicts of the representative attempts but that there might not be any representative attempts at all.

(2). *There might not be any representative attempts.* There are at least two ways in which it could turn out that there are no representative attempts. First, the agent might be unable to try. Consider a snail, for instance. There are no representative attempts by the snail to prove the Riemann Hypothesis. To imagine the snail even attempting to prove the Riemann Hypothesis requires imagining a significantly enhanced version of the snail. The same goes for me in a coma. It might be impossible for someone in a coma to attempt to run. If so, and if I am indeed in a coma, then there will be no representative attempts by me to run.

Second, it might be that all of an agent's attempts are done in circumstances that are, as compared to the present and actual circumstances, considerably less favorable. This brings me back, finally, to Nate and his factive abilities. We want to know whether Nate is presently and actually able to know that $h\&l$. Let us suppose that in the actual world, in which $h\&l$ is true, Nate has not, does not, and will not attempt to know that $h\&l$.[8]

representativeness it not illegitimate.

[7] There are similarities between my fourth counterfactual test and a proposed semantics for 'can' offer by Keith Lehrer. See his "'Can' in Theory and Practice: A Possible Worlds Analysis," in the Myles Brand and Douglas N. Watson (eds) *Action Theory* (Dodrecht: D. Reidel, 1976): 241-270.

[8] In some Nate Cases, modal space contains no representative attempts. In other Nate Cases, modal space contains some, but not enough, representative attempts. It simplifies the argument at this point to consider a Nate Case in which there no representative attempts.

There are possible worlds in which Nate does attempt to know that $h\&l$, of course. But every possible world in which Nate attempts to know that $h\&l$ is a world in which $h\&l$ is false, and a world in which $h\&l$ is false is considerably less favorable to Nate knowing that $h\&l$ than is a world in which $h\&l$ is true. Thus, while modal space contains many attempts by Nate to know that $h\&l$, none of those attempts are representative attempts.

The suspicion that abilities must be modally witnessed is, in my view, the result of combining a deep and important truth about abilities with an unfounded modal assumption. The deep and important truth is contained in the following conditional: if modal space contains enough representative attempts by S to ϕ, then S is able to ϕ iff a high enough proportion of those representative attempts are successes.[9,10]

(A few remarks about this conditional. **Q**: Why enough representative attempts rather than any? **A**: Because there are fluky successes and fluky failures. If there is just one representative attempt by S to ϕ and it is a fluky success, then 100% of S's representative attempts to ϕ are successes, but S is unable to ϕ. **Q**: What constitutes a high enough proportion? **A**: It seems to vary from ability to ability. If 30% of the agent's representative attempts to make a three-point basket are successes, then we conclude that the agent is able to make a three-point basket; on the other hand, if 50% of the agent's attempt to predict the outcome of a coin flip are successes, we do not conclude that the agent is able to predict the outcome of a coin flip.)

But to get from this conditional to the claim that abilities must be modally witnessed, we need a modal premise. We need the claim that modal space always contains enough representative attempts. And that is precisely what I deny. Sometimes modal space does not contain enough representative attempts. Moreover, the fact that modal space does not contain enough representative attempts does not decide the question of whether the agent has the ability: Nate is able to know that $h\&l$, but the snail is unable to

[9] Whether there are enough representative attempts it not a purely numerical matter. The representative attempts must also be sufficiently diverse.

[10] For more on proportionality, especially as it relates to dispositions, see David Manley and Ryan Wasserman, "On Linking Dispositions with Conditionals," *Mind* 117 (2008): 59-84, 76-82.

prove the Riemann Hypothesis.

I am inclined to think that the poss-ability principle is true when restricted to cases in which modal space contains enough representative attempts. There seems to be a necessary connection between abilities and successful representative attempts.[11] But the poss-ability principle admits exceptions when modal space does not contain enough representative attempts. There is no necessary connection between abilities and successful unrepresentative attempts: if all, or almost all, of an agent's attempts to ϕ are unrepresentative attempts, then the agent might be able to ϕ, even though all of the agent's attempts to ϕ are failures.

Sometimes, when modal space does not contain enough representative attempts, we can alleviate the problem by going vicarious. Take Nate. The problem is that none of Nate's attempts to know that $h\&l$ are done in circumstances that are as favorable to Nate knowing that $h\&l$ as are the present and actual circumstances. What we really want to know is whether an attempt by Nate *in the favorable circumstances that presently and actually obtain* would or might be successful, but looking to Nate's attempts, all of which are done in less favorable circumstances, does not provide us with that information. There is a second best, however. We can go vicarious. We can consider one or more agents who are enough like Nate—e.g. Nate's

[11] Are time travelers able to kill their younger selves? Here is one reason to think not. If S is able to ϕ, then a certain long run proportion of the representative attempts by S to ϕ will be successes. Now imagine an agent who lives for perpetuity. The agent is constantly traveling backwards in time and attempting to kill her younger self. Not only are there an infinite number of attempts by the agent to kill her younger self, there are an infinite number of representative attempts. Yet all of them are failures! This tests the bounds of credulity. If there are only one or two (or ten or fifty) representative attempts, then all of them might be fluky failures. But if there are an infinite number of representative attempts, and the attempts as diverse as can be, how could all of them be *fluky* failures? How could it be a fluke that every *possible* attempt at autoinfanticide ends in unsuccessful? This is the argument, more or less, that Kadri Vihvelin *op. cit.* presses against the claim that time travelers are able to kill their younger selves, and I have sympathy for it (although I do not find irresistible), which is why I believe that Nate Case pose a much more formidable threat to the poss-ability principle than do cases of attempted autoinfanticide.

classmates—who do indeed make attempts in the favorable circumstances that presently and actually obtain. The verdicts of the classmates' attempts are *vicariously* representative of whether Nate is presently and actually able to ϕ; they represent what would or might have happened if, *per impossibile*, Nate had attempted to know that $h\&l$ in the favorable circumstance that presently and actually obtain. (More concretely, they represent what would or might have happened if, *per impossibile*, Nate had attended the actual physics lecture that his classmates attended.) The vicarious attempts are successes—upon attending the lecture, Nate's classmates come to know that $h\&l$—and this suggests that Nate, too, is able to know that $h\&l$.

But as we will see when we consider Lonely Nate Cases, going vicarious cannot eliminate the problem entirely. Just as there might not be enough representative attempts, so there might be not be enough vicariously representative attempts. In such cases, the strategy of determining what an agent is able to do by looking to the verdicts of the (perhaps vicariously) representative attempts simply runs its course and peters out. There are not *always* enough representative attempts, so we cannot *always* determine whether the agent has the ability by looking to the verdicts of the agent's representative attempts.

Let me say a bit more about factive abilities. Factive abilities are a subset of object-dependent abilities. Object-dependent abilities are abilities to perform object-dependent actions, such as *visiting the Statue of Liberty*, *tasting vegemite*, or *kissing the Blarney Stone*. Factive abilities are the object-dependent abilities that depend on facts. Just as *visiting the Statue of Liberty* depends on the existence of the Statue of Liberty, so *knowing that* p depends on it being a fact that p (or in the language of existence: depends on the existence of the fact that p).

Whether a given circumstance is favorable to an agent performing an object-dependent action depends in large part on whether the object on which the action depends exists. For example, the times and worlds at which the Statue of Liberty exists are considerably more favorable to an agent *visiting the Statue of Liberty* than are the times and worlds at which the Statue of Liberty does not exist.

It is instructive to consider a temporal analog to a Nate Case—a *Jeremy Case*. Take a temporally fragile object, e.g. a human being, and consider an action and an ability that depends on that human being. Jeremy is at home on Monday. He considers whether to visit his grandmother, but decides against it. That evening, Jeremy's grandmother passes away. If Jeremy attempts to visit his grandmother on Tuesday (and suppose that he does), then his attempt will be a failure; his grandmother is not there to be visited. But Jeremy's attempt on Tuesday to visit his grandmother is not representative of whether Jeremy was, on Monday, able to visit his grandmother. The circumstances on Tuesday are, as compared to the circumstances on Monday, considerably less favorable. And if Tuesday sees Jeremy's very first attempt to visit his grandmother, then, although there are actual attempts by Jeremy to visit his grandmother, none of Jeremy's actual attempts to visit his grandmother are representative of whether he is, on Monday, able to visit his grandmother.

Jeremy's grandmother is a temporally fragile object. Nate Cases rely on the fact that there are actions and abilities that depend on modally fragile objects. Facts are modally fragile objects, so, when looking for counterexamples to the poss-ability principle, it is convenient to turn to factive abilities and factive actions. It is fact that $h\&l$, for example, but if the world were ever so slightly different in any of a host of ways, then it would not be a fact that $h\&l$.

If there are other modally fragile objects besides facts, then we can construct counterexamples to the poss-ability principle that make use of them. For instance, imagine a big rock in Siberia. Someone who is inclined by what Karen Bennett calls *bazillion thing-ism* thinks that there is an object that is just like the rock, the same size as the rock, the same shape as the rock, the same color as the rock, co-located with the rock, but different from the rock because it, unlike the rock, has all of its properties essentially.[12]

[12]Karen Bennett, "Spatio-temporal Coincidence and the Grounding Problem," *Philosophical Studies* 118 (2004): 339-371, 356-359. Also see Sarah-Jane Leslie, "Plenitude, Essence, and Paradox," *Philosophical Perspectives* (forthcoming); Mark Johnston, "Hylomorphism," *Journal of Philosophy* 103 (2006): 652-698; and Stephen Yablo, "Identity, Essence, and Indiscernibility," *Journal of Philosophy* 84 (1987): 293-314.

Suppose that nobody ever sees this big rock in Siberia, and thus that no one ever sees this modally fragile object. A question arises: is anyone—is Nate—able to see this modally fragile object?

I say that of course he is—it is not as if this modally fragile object is invisible! The object is as big as a rock, as brown as a rock, as conspicuous as a rock, and reflects photons just like a rock does—indeed, essentially so! It is metaphysically impossible for Nate to see this modally fragile object, sure. But it would be a mistake to infer from the fact that there is no possible world in which Nate *sees* this modally fragile object to the conclusion that there is no possible world in which Nate is *able to see* this modally fragile object. Nate is actually able to see this modally fragile object—Nate has what it takes to see big, brown, conspicuous, photon-reflectors, and he is not prevented from doing so—Nate simply does not exercise his ability.

It would be tendentious to rely on bazillion thing-ism in the present context, so I will not. Going forward, I focus on factive actions and factive abilities. I mention bazillion thing-ism only to emphasize that factive abilities are not necessarily the only abilities that are modally peculiar in this way. Whenever you have an object-dependent action—ϕing—that depends on a modally fragile object, modal space is thus divided into the few favorable worlds, in which the object exists, and the many unfavorable worlds, in which the object does not exist. If an agent—Nate—is in a favorable world, able to ϕ, does not ϕ, and if the favorable worlds taken together do not contain enough attempts by Nate to ϕ, then we have a Nate Case.

4 Variations on the Case

Once we have the blueprint for a Nate Case in mind, we can alter the case in a variety of ways. There are, for example, near-miss cases, in which Nate just barely fails to attain the status of knowledge. Here are two.

A diachronic case. — At one point in time, Nate knows that h. At another point in time, Nate knows that l. It just so happens, however, that Nate forgets that h before he learns that l. So there is no point in time when Nate knows both that h and that l, and hence no point in time when

Nate knows their conjunction. Nate does not (and hence necessarily does not) know that $h\&l$, but Nate seems to be able to know that $h\&l$.

A Gettier case. — Curious about the initial conditions and the laws of nature, Nate decides to ask his mother, who does indeed know that $h\&l$. However, by a curious twist of fate, aliens have randomly chosen one human being to abduct, just for a single night, and subject to experimentation, and Nate's mother is the unlucky Earthling. For the duration of the abduction, Nate's mother is replaced by an alien facsimile, who knows neither that h nor that l. Nate, who justifiedly takes the facsimile to be his mother, asks the facsimile, "What are the initial conditions and the laws of nature?" The facsimile, trying not to blow its cover, takes a wild guess and improbably gets it right, "$h\&l$." On the basis of the facsimile's bad testimony, Nate comes to the true and justified belief that $h\&l$, but Nate does not know that $h\&l$, as Nate fails the Gettier condition for knowledge. And because Nate never receives any further evidence, and never changes the basis of his belief, Nate never comes to know that $h\&l$. Nate does not (and hence necessarily does not) know that $h\&l$, but Nate seems to be able to know that $h\&l$.

There are Nate Cases in which Nate makes a representative attempt and suffers a fluky failure.[13] — Suppose that Nate does an Internet search: "What are the initial conditions of the universe and the laws of nature?" The result is successful; the true and complete specifications of both the initial conditions and the laws of nature are displayed on Nate's computer screen: "$h\&l_6$." Nate typically comes to know the true results of his Internet searches; his track record is good. But on this particular occasion, Nate makes an uncharacteristic blunder and misreads the screen, mistaking a '6' for a '0'. Thus Nate comes to the almost-true-but-false belief that $h\&l_0$, rather than the wholly true belief that $h\&l_6$. Nate does not (and hence necessarily does not) know that $h\&l_6$, but it stands to reason that Nate's attempt to know that $h\&l_6$ is a fluky failure, not a failure due to inability.

There are indeterministic Nate Cases, too. Consider the simplest. — Suppose that the universe is almost perfectly deterministic, but that at

[13]Such cases preempt the objection that Nate, in a Nate Case, is unable to make a representative attempt.

one critical juncture, long ago, the universe either zigged or zagged. As a matter of fact, the universe zigged. Let *h&l&zig* be the conjunction of the complete specification of the initial conditions, the complete specification of the almost deterministic laws of nature, and the claim that the universe zigged. Suppose that Nate has not, does not, and will not believe that *h&l&zig*. Still, Nate might be able to know that *h&l&zig*.

There might be incompatibilistic Nate Cases, as well. Incompatibilists maintain that nobody has any control over the initial conditions or the laws of nature. Let us say that a fact is *fixed* just if, by the lights of incompatibilism, nobody has any control over it. We may then distinguish necessitation from determination along the following lines. One fact *necessitates* another fact just if the one fact is incompossible with the negation of the other. One fact *determines* another fact just if the one fact is a fixed fact and is incompossible with the negation of the other. Heretofore I have focused on fixed facts, such as *h&l* or *h&l&zig*, which not only necessitate but also determine that Nate does not know them. Incompatibilists think that being determined not to ϕ is incompatible with being able to ϕ, so, if the fact that p determines that Nate does not know that p, then incompatibilists are unlikely to think that Nate is able to know that p. But my focus on fixed facts has been solely for the sake of convenience. All that we need is a fact that necessitates that Nate does not know it.[14] Take the true and complete description of the Milky Way. Call it m. If Nate does not know that m, then m necessitates that Nate does not know that m. It is therefore metaphysically impossible for Nate to know that m, as every world in which Nate believes that m is a world in which m is false. But m need not determine that Nate does not know that m, as m need not be a fixed fact. Or take the fact that the actual world is actual. I take the proposition *that the actual world is actual* to be a maximally contingent truth, i.e. actually true but such that it would be false if any actual truth was false. Some

[14] And everyone should think that necessitation is compatible with being able to do otherwise. After all, unexercised abilities are possible! If we could show that S is unable to ϕ by identifying a fact that necessitates that S does not ϕ, then the fact that S does not ϕ would imply that S is unable to ϕ.

agents are able to know that the actual world is actual, and others are not. The question arises whether an agent—Nate—might have the unexercised ability to know that the actual world is actual. Let the world be indeterministic. Let there be agent causal powers. Let the case be as concessive and compelling to incompatibilists as it can be made to be. Might Nate have the unexercised ability to know that the actual world is actual? I think so. Even when I don my incompatibilist hat, I think so. And if so, then we have a counterexample to the poss-ability principle, even by the lights of incompatibilism.

One variation deserves special attention. There are *Lonely Nate Cases*, in which Lonely Nate is the only being in all of modal space who is rightly positioned to perform a certain factive action. — Suppose that $h\&l$ is true, hence that determinism is true. In this universe, Lonely Nate is the only intelligent being. Lonely Nate has lived for millions of years and, in that time, developed all of modern science and mathematics. As it happens, Lonely Nate never comes to believe that $h\&l$, but this is an historical accident. (The reader may fill in the accidental details in any of a variety of ways: perhaps Lonely Nate never considers the question, or perhaps Lonely Nate makes an uncharacteristic blunder in his calculations and forgets to carry the one, etc.) We can supply Lonely Nate with all the requisite technology (perhaps Lonely Nate need only type the question into his supercomputer) and make Lonely Nate as able-minded as we like.

Lonely Nate is able to know that $h\&l$ iff Nate is able to know that $h\&l$. But Lonely Nate is an especially interesting case because it is not just metaphysically impossible *for Lonely Nate* to know that $h\&l$. It is metaphysically impossible *tout court* to know that $h\&l$—not a single being in all of modal space knows that $h\&l$. Lonely Nate is thus a counterexample to a principle that is even weaker than the poss-ability principle; namely, the impersonal poss-ability principle: If S is able to ϕ, then it is metaphysically possible to ϕ. Nate is able to do what it is metaphysically impossible for Nate to do, which is impressive. But Lonely Nate is able to do what it is metaphysically impossible *tout court* to do, which is more impressive still.

5 Consequences

If the poss-ability principle is false, then a number of interesting consequences follow. Let me mention three.

(1). *Modal theories of ability are false.* A core assumption of many who work in the philosophy of ability is that we can analyze, or anyway characterize, abilities in terms of metaphysically possible exercisings thereof. Almost all of the leading theories of ability are modal theories, where a *modal theory* is any theory of ability that takes the following form, either explicitly or implicitly: S is able to ϕ iff there is some possible world _____ in which S ϕs. If the poss-ability principle is false, then all modal theories of ability are false. Consider the two main modal theories of ability.

First, a counterfactual theory of ability. According to a counterfactual theory of ability, S is able to ϕ iff S would or might ϕ if S _____. The most famous counterfactual theory of ability is the conditional analysis of ability.[15] According to the conditional analysis of ability, S is able to ϕ iff S would or might ϕ if S tried to ϕ. On the standard Stalnaker-Lewis semantics for counterfactuals, 'S would or might ϕ if S tried to ϕ' is true iff some or all of the closest possible worlds in which S tries to ϕ are worlds in which S ϕs; similarly, 'S would or might ϕ if S _____' is true iff some or all of the closest possible worlds in which S does _____ are worlds in which S ϕs.

In Nate Cases, Nate is able to ϕ, even though there is no possible world in which Nate ϕs. In Lonely Nate Cases, Lonely Nate is able to ϕ, even though there is no possible world in which anyone ϕs. Nate Cases thus suggest that no counterfactual of the form, 'S would or might ϕ if S _____', is a necessary condition on S being able to ϕ.

Second, a restricted possibility analysis of ability.[16] There is a sense of

[15] The conditional analysis of ability was first proposed by G. E. Moore. See his "Free will," in *Ethics* (Oxford: Oxford University Press, 1912) 102-116.

[16] See Lewis *op. cit.*: p. 77: "To say that something can happen is to say that its happening is compossible with the facts. *Which* facts? That is determined, but sometimes not determined well enough, by context." Also see Angelika Kratzer, "What 'Must' and 'Can' Must and Can mean," *Linguistics and Philosophy* 1 (1977): 337-335; and Theodore Sider, *Writing the Book of the Worlds* (Oxford: Oxford University Press, 2011) 289-291.

'can' in which the following biconditional is uncontroversially true: S is able to ϕ iff S can (in the special sense) ϕ. This special sense of 'can' denotes a particular grade of possibility. To put names to ideas, let us call it *agentive possibility*. Hereafter we have three equivalent ways of speaking: S is able to ϕ iff S can (in the agentive sense) ϕ iff it is agentively possible for S to ϕ.

According to the restricted possibility analysis of ability, agentive possibility is a restricted grade of metaphysical possibility, where X-possibility is a restricted grade of Y-possibility iff to be X-possible is to be Y-compossible with a certain set of facts. The paradigm case is physical possibility, which is a restricted grade of metaphysical possibility; to be physically possible is to be metaphysically compossible with the nomic facts. If agentive possibility is a restricted grade of metaphysical possibility, then there is some (perhaps time-, individual-, and context-relative) set of facts—call them the *agentive facts*—such that to be agentively possible is to be metaphysically compossible with the agentive facts.

Much of the debate between compatibilists and incompatibilists seems to presuppose that agentive possibility is a restricted grade of metaphysical possibility. Alfred Mele, for example, echoing Peter van Inwagen's Second Argument for Incompatibilism,[17] says:

> Philosophers happy to talk in terms of possible worlds will say that an agent at a world W is [able] to A at t if and only if she (or a counterpart) A-s at t in some relevant possible world.... One way to see the disagreement between incompatibilists and compatibilists about determinism and being able to do otherwise is as a disagreement about what worlds are relevant. According to incompatibilists, all and only worlds with the same past and natural laws as W are relevant; they hold the past and the laws fixed. Compatibilists disagree.[18]

Mele portrays compatibilists and incompatibilists as agreeing from the onset that S is able to ϕ iff it is compossible with the agentive facts for S to ϕ,

[17] Peter Van Inwagen, *An Essay on Free Will* (Oxford: Oxford University Press, 1983) 86-93.

[18] Alfred Mele, "Agents' Abilities", *Noûs* (2003): 447-470, p.451.

and disagreeing merely about what the agentive facts are. (Are h or l, or both, agentive facts?)

There is a genuine disagreement in the vicinity in which Mele is gesturing. Compatibilists and incompatibilists genuinely disagree about which attempts are representative attempts. According to incompatibilists, an attempt by S to ϕ is representative of whether S is presently and actually able to ϕ only if the attempt occurs in a world with the same past and laws as the actual world; compatibilists disagree. But anyone who thinks that Nate is able to know that $h\&l$ certainly does not agree that S is able to ϕ iff it is compossible with the agentive facts for S to ϕ. After all, no matter what the agentive facts are, it is not compossible with them for Nate to know that $h\&l$!

If Nate is able to know that $h\&l$, then agentive possibility is not a restricted grade of metaphysical possibility; ability is not compossibility. There are metaphysical possibilities that are not agentively possible (e.g. me walking on water) and agentive possibilities that are not metaphysically possible (e.g. Nate knowing that $h\&l$).[19]

(2). *Modal arguments for incompatibilism, as they are currently formulated, fail.* The standard arguments for incompatibilism are modal arguments, and almost all of them rely either on the poss-ability principle itself or some principle that stands and falls together with the poss-ability principle. For example, in Peter van Inwagen's Consequence Argument, the premise that plays the part of the poss-ability principle is Rule Alpha. According to Rule Alpha: $\Box p \rightarrow \mathbf{N}p$. (In English: if it is metaphysically necessary that p, then nobody is able to render it false that p.) Many commentators characterize Rule Alpha as the one unassailable part of the Consequence

[19]The thesis that agentive possibility does not imply metaphysical possibility is similar to Leibniz's thesis that freedom requires only *per se* contingency. See Robert Merrihew Adams, *Leibniz: Determinist, Theist, Idealist* (Oxford: Oxford University Press, 1998) 10-22; Robert C. Sleigh Jr., "Leibniz on Freedom and Necessity: Critical Notice of Robert Adams, Leibniz: Determinist, Theist, and Idealist," *Philosophical Review* 108 (1999): 245-277; and Martin Lin, "Rationalism and Necessitarianism," *Noûs*, (forthcoming). Also see Susan Wolf, *Freedom Within Reason* (New York: Oxford University Press, 1990) 94-116.

Argument. Kadri Vihvelin says, "Rule Alpha is uncontroversial".[20] Alicia Finch and Ted A. Warfield say, "Rule Alpha is surely unobjectionable".[21] Peter van Inwagen says,

> I do not see how anyone could reject Rule (α). If (α) is invalid, then it could be that someone has a choice about what is necessarily true. Hardly anyone besides Descartes has been willing to concede such a capacity even to God. No one, so far as I know, has ever suggested that human beings could have a choice about what is necessarily true.[22]

We who accept that Nate is able to know that $h\&l$ do, however, reject Rule Alpha. More precisely, for any interpretation of '$\mathbf{N}$' on which it follows from $\mathbf{N}(S$ does not $\phi)$ that S is unable to ϕ, we claim that $\mathbf{N}p$ does not follow from $\Box p$. Nate Cases are counterexamples. It is necessarily true that Nate does not know that $h\&l$, yet Nate is able to know that $h\&l$.

If we reject the poss-ability principle, then we should reject the Consequence Argument. And the same goes for many other modal arguments for incompatibilism. Incompatibilists often attempt to show that determinism and the ability to do otherwise are incompatible by showing that it is impossible (given a certain set of facts) for agents in deterministic universes to do otherwise. But the fact that it is impossible (given a certain set of facts) for an agent to ϕ does not imply that the agent is unable to ϕ, so these arguments for incompatibilism, as they are currently formulated, fail.

To be clear, I do not think that rejecting the poss-ability principle should lead use to reject incompatibilism. (Indeed, I am inclined to think that compatibilists and incompatibilists should *agree* that the poss-ability principle is false.) It is an interesting and difficult question how arguments for incompatibilism should be reformulated in light of the failure of the poss-ability principle. (I have some ideas, but exploring them would take us too far

[20] Kadri Vihvelin, "Arguments for Incompatibilism," *Stanford Encyclopedia of Philosophy* (2011).

[21] Alicia Finch and Ted A. Warfield, "The Mind Argument and Libertarianism," *Mind* 107 (1998) 515-528, p. 517.

[22] van Inwagen *op. cit.* p. 96.

afield.) The point I want to make is simply this: that it is a consequence of rejecting the poss-ability principle that we must reassess and reformulate many of the main arguments offered in favor incompatibilism.

(3). *Metaphysical modality has fewer normative implications.* There are several putative connections between normativity and metaphysical modality, and just about all of them go by way of the poss-ability principle. If the poss-ability principle is false, then the connection between normativity and metaphysical modality is more attenuated than it is standardly taken to be. Consider willful Nate Cases, in which Nate's failure to ϕ is willful, deliberate, and morally weighty.

A willful case. — Suppose that $h\&l$ is true, hence that determinism is true. Nate receives a one-question true-or-false quiz. The question: "True-or-false: $h\&l$?" As it happens, Nate's father is up for the "PTA Father of the Year Award", which he both yearns for and deserves to win. Nate knows that if he fails the true-or-false quiz, then his father will lose the award. Out of nothing but willful and bitter teenage rebellion, sheer Schadenfreude, Nate intentionally fails the true-or-false quiz. Nate, who knows full well that $h\&l$, incorrectly answers that $h\&l$-is-false.

Correctly answering that $h\&l$-is-true is a factive action. It is an action that most of Nate's classmates perform, an action that is metaphysically impossible for Nate to perform, but an action that (in my view) Nate is able to perform, nonetheless. Notice that where ability and metaphysical possibility part ways, normativity goes with ability.

Nate ought (in the moral sense) to correctly answer that $h\&l$-is-true, and therefore Nate does as he ought not to do. There might be a true version of 'Ought' implies 'Can'. If so, then it is the principle that 'Ought' implies 'It is agentively possible that' (i.e. that 'Ought' implies 'Ability'). The principle that 'Ought' implies 'It is metaphysically possible that' is false. Willful Nate Cases are counterexamples. It is metaphysically impossible for Nate to correctly answer that $h\&l$-is-true, but Nate ought (in the moral sense) to correctly answer that $h\&l$-is-true.

Similarly, Nate is morally responsible and blameworthy for failing to correctly answer that $h\&l$-is-true. (Or equivalently: Nate is morally responsible

and blameworthy for incorrectly answering that $h\&l$-is-false.) Typically, if an agent ought to ϕ, but cannot, then the agent is thereby excused. But the fact that it is metaphysically impossible for Nate to correctly answer that $h\&l$-is-true does not excuse Nate from failing to do so. There might be a true version of the principle of possible action.[23] If so, then it is the principle of *agentively* possible action: an agent is morally responsible for failing to ϕ only if it is agentively possible for the agent to ϕ. The principle of *metaphysically* possible action is false. Willful Nate Cases are counterexamples. It is metaphysically impossible for Nate to correctly answer that $h\&l$-is-true, yet Nate is morally responsible and blameworthy for failing to correctly answer that $h\&l$-is-true.[24]

6 The Blankables

The *blankables* are a subclass of terms in English that end '-able' or '-ible', which include but are not limited to: 'visible', 'perceptible', 'discoverable', 'thinkable', 'sayable', 'provable', 'practicable', 'digestible', 'exercisable', 'fallible', 'tamable', and 'conquerable'. The blankables are to be analyzed in terms of ability.[25] A visible object is an object that can (in the agentive

[23]Peter van Inwagen, "Ability and Responsibility," *Philosophical Review*, 87 (1978): 201-224, 204.

[24]If we reject the poss-ability principle, then there is also an interesting consequence for deontic logic. There is a well-known attempt to reduce deontic logic to modal logic called the Andersonian-Kangerian reduction. Starting with the weak modal logic **K**, we can add a propositional content d, the claim that all normative demands are met, and the axoim A3: $\Diamond d$. The resultant logic is called $\mathbf{K}d$, and it can be shown that all of the theorems of standard deontic logic are contained in $\mathbf{K}d$. On the Andesonian-Kangerian reduction, permission is a restricted grade of possibility: p is permissible iff p is compossible with d; obligation is a hypothetical grade of necessity: p is obligatory iff d necessitates p.

Nate Cases make trouble for the Andersonian-Kangerian reduction in the obvious way. The attempted reduction relies on the viability of $\mathbf{K}d$, but Nate Cases show that A3 is false. Some normative demands are impossible to meet. Indeed, neither the claim that p is permissible, nor the claim that p is obligatory, imply that p is possible. Thus, the Andersonian-Kangerian reducton of deontic logic to modal logic fails.

[25]There is another class of words that end '-able' or '-ible' that should not be confused with the blankables. 'Admirable', 'desirable', 'detestable', 'valuable', and 'enviable',

sense) be seen. A knowable object is an object that can (in the agentive sense) be known.

It is standard practice is philosophy to regiment the blankables in terms of metaphysical possibility. The claim that x is visible is regimented in symbols as '$\Diamond Sx$', i.e. that it is metaphysically possible that x be seen. The claim that p is knowable is regimented in symbols as '$\Diamond Kp$', i.e. that it is metaphysically possible that p be known. If the poss-ability principle was true, then the modal regimentation of the blankables might be legitimate. But I think the poss-ability principle is false, and that the modal regimentation is not just illegitimate but also the source of some philosophical confusion.

Start with exercisability. Some philosophers might object to my thesis that the poss-ability principle is false on the grounds that abilities must be exercisable. But that objection misses the mark, for I agree that abilities must be exercisable. The crucial point is that I insist that exercisability is a matter not of metaphysical possibility but of ability. Something might be exercisable, even though it is necessarily unexercised; Nate's ability to know that $h\&l$ is a case in point.

In general, the inference from the fact that something is necessarily unblanked to the conclusion that the thing is unblankable strikes me as dubious. I suspect that some perceptible things are necessarily unperceived; I suspect that some visible things are necessarily unseen (and I gave a potential example in §2); I suspect that some tangible things are necessary untouched; and I suspect that some knowable things are necessarily unknown. I want to spend the last part of this essay looking at knowability from a slightly different angle.

Consider the anti-realist's contention that knowability is a constraint on truth: if p is true, then p is knowable. There is a famous argument against the knowability constraint, first published by Fitch.[26] The argument begins

are—on their most natural readings—of a kind with 'trustworthy' and 'newsworthy': to be analyzed not in terms of ability, but in terms of fittingness or worthiness. To be admirable is to be fitting to admire. To be enviable is to be fitting to envy.

[26]Frederic Fitch, "A Logical Analysis of Some Value Concepts," *Journal of Symbolic Logic* 28 (1963): 135-142.

by regimenting knowability as '$\Diamond K$' and proceeds as follows:

(1)	$\forall p\ (p \rightarrow \Diamond Kp)$	premise, the knowability constraint
(2)	$\exists q\ (q\ \&\ \neg Kq)$	premise, some truth is unknown
(3)	$(q\ \&\ \neg Kq) \rightarrow \Diamond K(q\ \&\ \neg Kq)$	from (3), by the knowability constraint
(4)	$\Diamond K(q\ \&\ \neg Kq)$	from (2) and (3), by sentential logic
(5)	$\Diamond(Kq\ \&\ K\neg Kq)$	from (4), by $\Box(K(p\ \&\ q) \rightarrow (Kp\ \&\ Kq))$
(6)	$\Diamond(Kq\ \&\ \neg Kq)$	from (5), by the factivity of knowledge

But (6) is false; contradictions are not possibly true.

How should an anti-realist respond to Fitch? I suggest that the anti-realist reject the modal regimentation of knowability.

It is helpful in this context to consider two cases. First, Lonely Nate. Lonely Nate is the only being in all of modal space who is rightly positioned to know that $h\&l$, yet Lonely Nate fails to know that $h\&l$. In symbols, $\neg\Diamond K(h\&l)$. But it stands to reason $h\&l$ is not the sort of claim that the anti-realist intends to target. It is impossible to know that $h\&l$, but $h\&l$ is not (in the anti-realist's preferred sense of the term) unknowable. Second, conjunction. Take the conjunction of the (atomic) propositions that are knowable by the anti-realist's lights. Suppose that this conjunction is not actually believed. Then it may well be impossible for this conjunction to be known. After all, this conjunction contains many facts about which mental and physical states do and do not obtain, but different mental and/or physical states would have to obtain in order for this conjunction to be believed. In other words, it might be the case that this conjunction is believed only in worlds in which it is false. Nevertheless, this conjunction, although it may be necessarily unknown, is not (in the anti-realist's preferred sense of the term) unknowable. Anti-realists do and should believe that knowability is closed under conjunction.

I think that the anti-realist who is attracted to the knowability constraint on truth should deny the poss-ability principle, repudiate the modal regimentation of knowability, insist that there are knowable but necessarily

unknown truths, and thus stop Fitch's Paradox before it starts.[27] If (4) is supposed to mean, 'It is knowable that (q & $\neg Kq$)', then the inference from (4) to (5) is invalid. The fact that a certain conjunction is knowable does not imply that there is a metaphysical possibility in which each of the conjuncts is known.

A similar move resolves a knowability paradox that David Chalmers has recently put forward.[28] Chalmers introduces his paradox as follows.

> It is widely believed that for all p, or at least all entertainable p, it is knowable a priori that (p iff actually p). It is even more widely believed that for all such p, it is knowable that (p iff actually p). There is a simple argument against these claims from four antecedently pausible assumptions.... [Let] 'A', 'E', 'K', '$\Box$', '$\Diamond$' stand for 'Actually', 'Someone entertains', 'Someone knows', 'Necessarily' and 'Possibly', while '$\rightarrow$' and '$\leftrightarrow$' are the material conditional and biconditional. In addition, q is any (entertainable and expressible) proposition that no-one actually entertains, while r is $\neg Eq$, the proposition that no-one entertains q.
>
> (1) Ar
>
> (2) $Ar \rightarrow \Box Ar$
>
> (3) $\Box(K((r \leftrightarrow Ar) \rightarrow (r \leftrightarrow Ar))$
>
> (4) $\Box(r \rightarrow \neg K(r \leftrightarrow Ar))$
>
> ———
>
> (5) $\neg\Diamond K(r \leftrightarrow Ar)$

[27] This strategy for responding to Fitch's Paradox appears in Michael Fara, "Knowability and the Capacity to Know," *Synthese* 117 (2010): 843-865. Fara denies the poss-ability principle—or, strictly speaking, denies the *poss-capacity principle*: if S is capable of ϕing, then it is metaphysically possible for S to ϕ—and he regards cases of attempted autoinfanticide as counterexamples. Fara does not offer counterexamples along the lines of Nate Cases, does not rebut Rule Alpha, and does not deny the putative connections between normativity and metaphysical modality. My response to Chalmers below is an extension of Fara's reply to Fitch.

[28] David J. Chalmers, "Actuality and Knowability," *Analysis* 71 (2011): 411-419, pp. 411-412.

> ...The conclusion follows from the premises by classical logic and weak modal logic **K**.... From (3) and (4), one can derive $\Box(K(r \leftrightarrow Ar) \rightarrow (r \leftrightarrow Ar)\ \&\ \neg r)$. From (1) and (2), one can derive $\Box(K(r \leftrightarrow Ar) \rightarrow Ar)$. From these two claims, one can derive $\Box(K(r \leftrightarrow Ar) \rightarrow (r \leftrightarrow \neg r))$, from which the conclusion follows.

The intuitive argument goes like this. Every world is either an r-world or an $\neg r$-world. There is no r-world in which someone knows that $(r \leftrightarrow Ar)$, as there is no r-world in which someone entertains that $(r \leftrightarrow Ar)$. One must entertain that $(r \leftrightarrow Ar)$ in order to know that $(r \leftrightarrow Ar)$, and one must entertain that q in order to entertain that $(r \leftrightarrow Ar)$, but the r-worlds are precisely the worlds in which no-one entertains that q. At the same time, there is no $\neg r$-world in which someone knows that $(r \leftrightarrow Ar)$, as $(r \leftrightarrow Ar)$ is false at $\neg r$-worlds, and a proposition cannot be known at a world at which it is false.

Has Chalmers shown that there are entertainable truths of the form $(p \leftrightarrow Ap)$ that are unknowable? I think not. How do I respond to Chalmers's argument? In the form of a dilemma.

Either '$\Diamond$' and '$\Box$' are metaphysical possibility and necessity or agentive possibility and necessity. If they are metaphysical possibility and necessity, then the proof is sound, the conclusion is true, but the paradox is falsely advertised. If '$\Diamond$' is metaphysical possibility, then $\neg\Diamond K(r \leftrightarrow Ar)$ says that it is not metaphysically possible for someone to know that $(r \leftrightarrow Ar)$—which is true. But the fact that $(r \leftrightarrow Ar)$ is necessarily unknown does not imply that $(r \leftrightarrow Ar)$ is unknowable. There are knowable truths that are necessarily unknown, and I submit that $(r \leftrightarrow Ar)$ is one such.

If Chalmers's argument is to have the paradoxical conclusion it advertises, then '$\Diamond$' and '$\Box$' must be agentive possibility and necessity. If '$\Diamond$' is agentive possibility, then $\neg\Diamond K(r \leftrightarrow Ar)$ says that it is not agentively possible for someone to know that $(r \leftrightarrow Ar)$. (Remember: it is agentively possible for someone to know that $(r \leftrightarrow Ar)$ iff someone is able to know that $(r \leftrightarrow Ar)$ iff it is knowable that $(r \leftrightarrow Ar)$.) But if '$\Diamond$' and '$\Box$' are agentive possibility and necessity, then premise (2) is false. The fact that no-one

actually entertains that q does not imply that no one is *able* to actually entertain that q. We are supposed to accept (2) because we are supposed accept a principle of which (2) is an instance, namely, $(\forall p)\ Ap \rightarrow \Box Ap$. But if '$\Box$' means 'It is agentively necessary that', then $(\forall p)\ Ap \rightarrow \Box Ap$ is the principle of actualized fatalism; it says that one is able to actually do only what one actually does. And just as the principle of fatalism—i.e. $(\forall p)$ $p \rightarrow \Box p$—is refuted by unexercised abilities, so the principle of actualized fatalism is refuted by unexercised actualized abilities. If I reach a fork in the road, then I am able to actually go left and able to actually go right, even though I actually go only one way or the other. The fact that I do not actually entertain that q does not imply that I am not *able* to actually entertain that q, for my ability to actually entertain that q might be unexercised.[29]

In the end, then, Chalmers's argument is not a paradox but a proof. It is, though, I thinka very interesting proof. For Chalmers has shown that there are flesh and blood counterexamples to the adage that nobody is able to do the impossible. Or, putting the point another way, Chalmers has proved that you and I are like Lonely Nate. You and I are both able to know $(r \leftrightarrow Ar)$, as is anyone who is able to entertain the proposition, even though it is metaphysically impossible (not just for you or for me, but *tout court*) to know that $(r \leftrightarrow Ar)$. You and I are able to do what it is metaphysically impossible to do. You and I are able to do something that not a single being in all of modal space manages to do. Fancy that.[30]

[29] Chalmers considers the objection that I want to push, namely to accept the conclusion of his argument but reject the paradox. "One... strategy appeals to agentive possibility... holding that a proposition is knowable when it is agentively possible for someone to know it, while denying that agentive possibility entails metaphysical possibility. This position faces an obvious challenge, however, in that the original argument might be reformulated in terms of agentive possibility, and the case for the four key premises remains strong when read this way." I respond, as I do in the text, that the case for premise (2) does not remain strong when read in terms of agentive possibility.

[30] This paper was presented at the Princeton Philosophical Society in the fall of 2011 and the Princeton Dissertation Seminar in the spring of 2012. The feedback from those presentations lead to several improvements. A special thank you to Nathan Gadd, Elizabeth Harman, Andrew Huddleston, Mark Johnston, Boris Kment, Gideon Rosen, and Joe Rachiele.