

**WRETCHED SUBTERFUGE:
A DEFENSE OF THE COMPATIBILISM OF FREEDOM AND NATURAL CAUSATION**

The John Dewey Lecture
Pacific Division Meetings of the American Philosophical Association
March, 2010

John Perry
University of California Riverside
Stanford University

INTRODUCTION

This afternoon I will defend Hume's 'reconciling project' with respect to the issue of freedom and determinism, or, as he put it, 'liberty and necessity':

...to proceed in this reconciling project with regard to the question of liberty and necessity; the most contentious question of metaphysics, the most contentious science; it will not require many words to prove, that all mankind have ever agreed in the doctrine of liberty as well as in that of necessity, and that the whole dispute...has been hitherto merely verbal. For what is meant by liberty, when applied to voluntary actions? We cannot surely mean that actions have so little connection with motives, inclinations, and circumstances, that one does not follow with a certain degree of uniformity from the other, and that one affords no inference by which we can conclude the existence of the other. For these are plain and acknowledged matters of fact. By liberty, then, we can only mean a power of acting or not acting, according to the determinations of the will; that is, if we choose to remain at rest, we may; if we choose to move, we also may. Now this hypothetical liberty

is universally allowed to belong to everyone who is not a prisoner and in chains. Here, then, is no subject of dispute.¹

Hume's strategy, basically shared with Hobbes² and Locke³, is now called 'compatibilism'. A long line of distinguished philosophers agree with Immanuel Kant, that compatibilism is a 'wretched subterfuge':

It is a wretched subterfuge to seek to evade [the problem of determinism and freedom] by saying that . . . the actions of the human being, although they are necessary by their determining grounds which preceded them in time, are yet called free because the actions are caused from within, by representations produced by our own powers, whereby desires are evoked on occasion of circumstances and hence actions are produced at our own discretion.⁴

I want to discuss freedom, not free will. As I use the term, 'freedom' means having the ability to do other than one actually does. I'll borrow Carl Ginet's way of posing the issue.⁵ Suppose X is deliberating about whether to do A or to refrain from doing A. Such a person may think, 'I can do A, and I also can refrain from doing A'. Suppose X does in fact refrain from doing A, because on balance he wants to and chooses to; there is no coercion, in the ordinary sense of the word, of any sort. Suppose further that this choice is naturally caused, in the sense that it is determined by the laws of nature, given the way things were in the more or less remote past, to the same extent as

¹ David Hume, *Inquiry Concerning Human Understanding*, many editions, chapter 8.

² See Thomas Hobbes and John Bramhall, *Hobbes and Bramhall on Liberty and Necessity*, ed. Vere Chappell, Cambridge: Cambridge University Press, 1990.

³ John Locke, *An Essay Concerning Human Understanding*, many editions, Book II, Chapter XXI, 'Of Power'.

⁴ Immanuel Kant, *Critique of Practical Reason*, Book I, chapter 3.

⁵ From his defense of incompatibilism, in his excellent book *On Action*, Cambridge: Cambridge University Press, 1990.

other events in nature. Does it follow that X's judgment, that he could do *A*, was false? The incompatibilist says "yes"; Hume and I say "No".

I believe freedom is what I shall call a *natural* concept. By this I mean that it gets at a property of importance in human life, that humans came to recognize and take into account early in their history, long before language and philosophy rendered life the charming affair it has become. In contrast, I believe that free-will is not a natural concept in this sense, but a theoretical concept from Christianity infused with Greek philosophy, due particularly to Saint Augustine. He thought that Adam had free will, but since then humans have not had it, in that we do not have the power to do the right thing without the grace of God.

Pelagius thought, in contrast, that at least some humans some of the time retain the power to choose right from wrong, of their own "free will", and get rewarded for making the right choices. Augustine won the battle, and Pelagius's view became heresy, but perhaps Pelagius won the war, in that his view is probably is closer to the view of most contemporary Christians than Augustine's. Both Pelagius and Augustine were working with a dualist concept of the mind and the self, as something immaterial that works on different principles than ordinary physical things. This view concept of free will survives in the thinking of many philosophers, who do not all see themselves as Christians or dualists, as an analysis of freedom, or an important subcategory of freedom --- "contra-causal freedom," perhaps.⁶

To borrow an example from Manuel Vargas,⁷ consider the ancient view that water is one of the four basic elements. Our concept of water, like our concept of freedom, is a natural concept. The importance of water to human life can make us pretty confident that humans recognized water, dealt with

⁶ See C.A. Campbell, "Is 'Freewill' a Pseudo-Problem?" in [Mind](#), LX(240), October 1951: 446-465.

⁷ In "Revisionism," in *Four Views on Free Will*, by John Martin Fischer, Derek Pereboom, Robert Kane, Oxford: Blackwell Publishing, 2007.

water, and had ideas and concepts for dealing with water before they had language and words for water, and long before some philosopher came up with the false view that it was one of the four basic elements. That was a not terribly implausible theory about water, but it turned out to be false, and that's what I think about free-will vis-à-vis freedom. Freedom is a natural concept of an important property; free-will is a theory about that property, that it involves some kind of special causation, that doesn't seem very plausible now, given what we know, or think we know. Of course, we don't continue to teach people that water is one of the four basic elements in Sunday School and Freshman Thought and such places. At any rate, the theory that freedom involves free-will, the sort of contra-causal freedom that Pelagius and Augustine agreed that Adam had, has no appeal as far as I can see on the evidence available to us. I have no wish to claim that free will is compatible with natural causation, or that in the end it makes much sense at all.

So, to repeat, I am defending the compatibilism of freedom with natural causation. And by freedom I mean that natural property, of being able to do one thing or not do it, that humans have recognized that they sometimes have and sometimes don't have, since long before there were words to describe it or philosophers to misdescribe it.

As I said, I borrow Ginet's way of posing the issue, in terms of whether the thoughts we think using the word 'can' are true. I shall assume that 'can' connects with words like 'free' and 'responsible' in the usual ways. If we do *A* freely, then we could have refrained from doing *A*. If we cannot do *A*, that means that we are *not* responsible for *not* doing *A*. We may be responsible for not *trying* to *A*, if we think we can do *A*, because we are mistaken about our situation or our abilities --- because a door is locked that we think is unlocked, or a muscle is disabled that we don't know about, or a Frankfurtian intervener will stop us if we try.⁸ And even if we can't do *A*, we may deserve credit or

⁸ Harry G. Frankfurt, "Alternative Possibilities and Moral Responsibility," in Watson, op. cit.: 167-176.

blame for not being able to do *A*. If a heroine bites off her tongue before being tortured, so that she will not be able to disclose the secret information, perhaps she is then not responsible for not disclosing the information, and deserves no credit for this, but she is responsible for, and deserves credit for, biting off her tongue so that she couldn't disclose the information.

Frankfurt cases show, it seems, that we might do something, and be responsible for doing it, although we couldn't have done anything else, because had we decided to do something else, or even started to decide to, there would have been an intervention. So freedom is not necessary for responsibility. Be that as it may, most cases are not Frankfurt cases; paradigmatically, at least, when we are responsible for something, we could have refrained from doing it.

In the use we are interested in, the word 'can' serves to build a predicate from a verb phrase. If I say "I can move my arm," we have a predicate, 'can move my arm' built out of the verb phrase, 'move my arm'. The predicate expresses a property that a person has at time or doesn't have at that time. High school grammar tells us there are four main kinds of properties we thus express, competence, opportunity, permission and possibility:

- I can ride a horse. (competence)
- We can stay with my brother when we are in Paris. (opportunity)
- She can stay out after 10 PM. (permission)
- Any child can grow up to be president. (possibility)⁹

⁹ There is also a common use of 'can' in making requests:

Can you hand me the stapler? (request)

The request use is an implicature, perhaps a generalized implicature, in Grice's sense.⁹ I am asking if you have the ability to do something, in order to convey that I'd like you to do it.

It's important to remember that that the word 'could' serves as both the past tense of 'can', and as the subjunctive. If I say, "I didn't sleep late this morning, but I could have," I am just saying what I would have said this morning with "I can sleep late". Now consider this Brussels sprout here on the podium. I believe it is true that I can eat it a moment from now; more about this later. It is not true that I can drink a glass of whiskey, since there is no glass of whiskey here to drink. But it is true that I could drink a glass of whiskey, if there were one on the podium. I can't drink a glass of whiskey now, because there isn't one to drink, but I could drink one, if there were one here to drink. I have the competence, but not the opportunity. Later, I can say that I couldn't drink a glass of whiskey --- past tense --- , but I could have if there had been one there --- subjunctive.

I believe that the use of the word 'can' that is relevant to answering the question that Ginet asks, gets at a property that people have at a given time that is a matter of competence and opportunity. I can eat this Brussels sprout in a moment, because I have the competence to move my arms and hands at mouth in certain ways, and I have the opportunity to do so. Moving my arms and hands and mouth in that way --- the competence --- in the present circumstances --- there being a Brussels sprout on the podium in front of me -- - is a *way of* eating the a Brussels sprout. So I can do it.

These 'can' properties, these abilities, play an extremely important role in human life. They are important in deliberation. In deciding what to do, it's very helpful to know what one can and can't do. Hilary Bok has done a wonderful job illuminating the phenomenology of deliberation, and the importance of these properties in it, and if I had time I would quote her at length.¹⁰ But clearly, if I am deliberating about what to do, a natural starting point is to consider what I have the competence and opportunity to do, that is,

¹⁰ Hilary Bok, "Freedom and Practical Reason," in Gary Watson, editor, *Free Will*, Oxford: Oxford University Press, 2003: 130-166

to consider my options. The property is also important in predicting what other people will do. They won't do what they can't do, that is, what they lack the competence to do, or lack the opportunity to do, or both. They *also* won't do what they have no reason to do, but we are often in a position to figure out what they can and cannot do, when we have little information about what they have reason to do, and vice versa. They are different issues. Finally, if a person doesn't do something we wanted or expected them to do, how we think of them will be different if we find out that they didn't do it because they didn't have the competence or opportunity, than if we find out that the reasons we hoped would motivate them did not.

If my wife tells me to pick up a carton of lite milk at the store --- not whole, not skim, but lite --- and I arrive home with a carton of skim, she may be irritated, if she thinks that I simply didn't take her preferences seriously, or wasn't listening. But if I explain that the store was out of lite milk, so I didn't have an opportunity to get lite milk, I *couldn't* get lite milk, that should get me off the hook.

So my thesis is threefold: first, that 'can', as we use it in deliberation --- the Ginet 'can' --- stands for a property, freedom, involving competence and opportunity; that this is an important property in human life; it makes sense that we would have a word for it, and use it in deliberating about what to do, and in explaining and predicting what other people have done or will do, and in adjusting our feelings about what they have and haven't done. Second, if this is what the can-properties are, then it turns out that we can do things that we will not do, and could have done things we did not do, even if what we do is caused in the same way that other events in the natural world are. That is, an act can be free, in the sense that we could have done otherwise, even if the state of the relevant part of the universe at some time previous to our doing it or not doing it and the laws of nature settle that we will do it. Third, incompatibilist arguments to the contrary derive from false things philosophers believe about this property.

Here is my plan. First, I will develop a version of compatibilism. I argue that there is a property that X is thinking about when he thinks "I can A", and that he can be correct, that he has this property, even if the end he refrains from A-ing. I explain what my candidate for that property is. Then I take a critical look at the classical version consequence argument. Then I focus on Ginet's version and in particular his principle of the fixity of the given past. Finally, I will see if I can distinguish my view from Manuel Vargas', which he calls revisionism.

One more preliminary. I distinguish between determinism and what I call "natural causation." Natural causation is the view that human actions are part of the natural world, subject to laws of nature, in the same way that other complex events are, and, like other such events, are generally the results of causal chains that extend into the more or less remote past. If an event that did happen, had not happened, that would be because something about the past was different. Determinism is the view that the laws of nature admit of no exceptions and determine every aspect of every event down to any level of detail. If determinism is true, if an event that did happen, had not happened, something about the past would have been different, stretching back to the Big Bang, or God saying "Let it Be", or perhaps just infinitely.

I don't believe in determinism. This is partly because of quantum physics seems to indicate that the world is not deterministic, although admittedly the correct interpretation of quantum physics remains a live issue, involving many things I don't understand. It is partly because I think determinism has the smell of a religious or ideological fantasy. It is partly because it seems to me a very depressing view. And it is partly because various philosophers I respect, who know a lot about the issues, like Pat Suppes, Nancy Cartwright and John Dupre, don't believe in it. But, perhaps in contrast to some of them, I think that a compatibilist account of freedom is still required, if we are to have freedom, because I can't see that the looseness in the universe quantum physics seems to provide has anything to do with

freedom. I see no reason to think that paths that fork off from the actual sequence of events that correspond to events that could have happened but didn't because of quantum randomness, correspond to the paths that fork off from the actual sequence of events that correspond to actions that someone could have chosen to perform but did not. I don't think the Libertarian alternatives to the compatibilist account, such as Chisholm's agent causation¹¹, and Kane's theory,¹² that depend on such correspondence, are viable. Agent causation doesn't make much sense; Kane's account seems coherent, but I can see no reason to suppose that it is correct about what we are getting at when we think we can do something. And, if the Consequence Argument is a good argument against the compatibility of freedom and determinism, I think it would provide the basis of a good argument against the compatibility of freedom and natural causation. So, it seems to me, we need to defend compatibilism, and address the consequence argument, even if we have doubts about determinism.

Still, I assume that if freedom is shown to be compatible with determinism, it will also have been proven compatible with mere natural causation. Determinism is the focus of the literature I discuss. So from now on I will mainly focus on the compatibility of freedom with determinism.

OPTIONS FOR THE COMPATIBILIST

Suppose I am in some condition φ , and it is a true generalization about humans that if they are in condition φ , then they will not do some action A . Then it follows that I won't do A , but it doesn't follow that I can't do A . All that follows is that if I did do it, the generalization wouldn't be true.

¹¹ Roderick M. Chisholm, "Human Freedom and the Self," in Watson, op.cit.: 26-37.

¹² Robert Kane, "Responsibility, Luck and Chance: Reflections on Free Will and Indeterminism" in Watson, op.cit.: 299-321.

Now suppose that we add that the generalization is a law of nature. It still follows that I won't do *A*. But does it follow, from those strengthened premises, that I *can't* do *A*? The compatibilist says "no," the incompatibilist says "yes".

Hume's attack on incompatibilism has two prongs. The first is what I call a *weak theory of laws*; the laws of nature are simply true universal generalizations about what happens; their necessity is subjective, a matter of our internalizing the regularities the generalizations encode, and forming our expectations in accordance with them. So by adding that our generalization is a law of nature, we haven't really added anything to the picture, that would show that I can't do *A*. I don't accept this account of laws of nature, and so will not defend this prong.¹³

The second prong is a *weak theory of ability*. It doesn't amount to as much as you might have thought to have an ability to *A* --- to be in a position so that you can truly think or say, "I can *A*." So it requires *more* to show that one doesn't have such an ability, than one might have thought. Even if going from a true generalization to a law of nature adds *something* to the argument, it doesn't add enough to get us to the conclusion that I *can't* do *A*.

This prong reflects the aspect of Hume's thinking put forward in the quote above, essentially in the sentence,

¹³ While I agree with David Lewis's analysis of an ambiguity about 'render false' in van Inwagen's version of the Consequence Argument, ("Are We Free To Break the Laws?" in Watson, op.cit.: 122-129) I do not agree with his conclusion that if, in a deterministic universe, if someone had done something they didn't in fact do, they would have done it because of a local miracle that occurred shortly before. I believe Lewis holds this view because it fits best with various other views he has, including his view about counterfactuals, and the direction of time. But ultimately it rests on his somewhat Humean view of laws; Lewis' theory is a sophisticated version of Hume's first prong. I won't have time to say more about Lewis.

By liberty, then, we can only mean a power of acting or not acting, according to the determinations of the will; that is, if we choose to remain at rest, we may; if we choose to move, we also may.

Here is an analogy, that may help make the strategy clear. Consider financial ability. Suppose Elwood is considering whether to buy a Porsche convertible or a Ford compact. Suppose all of Elwood's funds that could be used to buy a car are in a certain bank account, into which Elwood's uncle has made a generous deposit. Although Elwood doesn't know exactly how much has been deposited, he thinks it is plenty to do whatever he wants. If the bank account contains more than \$100,000, Elwood has the financial ability to buy a Porsche; otherwise he does not. If it contains only, say, \$15,000 he has the financial ability to buy a small Ford, but not to buy a Porsche.¹⁴

Now we have two issues, clearly different. One is which car Elwood prefers. This issue will be decided by a process of deliberation on Elwood's part. The process involves his getting information from the external world; looking at cars, reading brochures. But, eventually, it is the unfolding of events in Elwood's brain, thinking through his what he really wants, that will determine his preferences. Then there is the issue of his financial abilities; of which car he can afford. This is not decided by events in his brain, but by the state of his bank account, something that probably supervenes on the state of the bank's computer, blocks away from Elwood's brain.

Suppose there is \$150,000 in the bank account. Then Elwood has the financial ability to buy a Porsche, quite independently of whether he in the end prefers a Porsche or a Ford. If he decides to buy a Ford, then he won't buy a Porsche. Nothing is forcing him to buy a Porsche. It's up to him. But

¹⁴ I gather from the internet that \$100,000 may not be enough for Elwood to get the sort of Porsche I had in mind for him, although it does seem to be enough to buy some sort of Porsche.

his decision not to buy a Porsche doesn't mean that he *can't* buy a Porsche; at least in this financial use of the term `can', he clearly can.

This example isn't supposed to convert any incompatibilist, but merely to help make clear the structure of the compatibilist's picture. (By `the compatibilist' in contexts like this, I mean my version of the compatibilist.) Abilities are one thing, and that's what the ideas of `can' and `cannot' as they occur in Elwood's thinking have to do with. Preferences, desires, and the like are another, and that's what will determine whether Elwood tries to buy the Porsche or the Ford.

Consider this Brussels sprout and I. The issue of whether I want to eat the Brussels sprout is one thing. The issue of whether I *can* eat it is another. Here the issue is not determined by the state of the computer in the bank down the street, but facts about my situation and physical dexterity. Is there a Brussels sprout in front of me to be eaten? Yes. Do I have the dexterity to reach the sprout and transport it to my mouth? To chew it up and swallow it? Yes. Then I can eat it, whether I want to or not. Facts about his structure of my preferences and beliefs, the factors that will determine the outcome of my deliberation about the Brussels sprout, are one thing; facts about the presence of a Brussels sprout, the length of my arm, and my coordination are something else.

RAW ABILITIES

There are two basic parts to my analysis, *raw abilities* and *refined abilities*. If one can do *A*, one has the raw ability to do *A*. I now have the raw ability to eat this Brussels sprout. However, I can't eat this Brussels sprout and still have a Brussels sprout as a prop for the rest of the lecture. So, in my terminology, I don't have the refined ability to both eat the Brussels sprout and also have one available for a prop. Refined abilities are important in accounting for some of our thinking about what we can and cannot do. This afternoon, however, I will only have time to discuss raw abilities.

The concept of a raw ability depends on that of a *competence to execute a type of movement*. Having such a competence is a property of a person, usually acquired, usually more or less enduring, but often not lasting forever. One's competences usually depend on anatomy and training, and can be lost by falling out of shape, not getting enough practice, or permanent disability. Competences can also be temporarily impaired, through a temporary injury, or external constraints, like chains. I have the competence right now to move my arm and hand in a way that will constitute reaching out and grabbing this Brussels sprout, because my arms are long enough and I have learned how to pick things up that I can reach. This competence is unimpaired at the present moment, because I haven't been injured, my arms are not chained, and so on. I may eventually lose it, because of arthritis or some other problem.

An agent has *the raw ability* at to *A* if:

- (i) There is a type of movement *M*, that the agent has the unimpaired competence to execute;
- (ii) The agent is in a certain type of circumstance *C*;
- (iii) Executing *M* in circumstances of type *C* is a *way of A-ing*.

I have the raw ability to pick up this Brussels sprout because I have an unimpaired competence to move my arm and hand in the requisite way, and I am in a circumstance in which exercising that competence will constitute picking up the Brussels sprout. I can do, what I have the raw abilities to do, *whether I want to or not*.

Right now, standing here at this podium, I have innumerable raw abilities. I can eat this Brussels sprout. I can jam this pen up my nose far enough to make me bleed in a disgusting way all over the podium. I could get undressed and sing "Oh Canada". I could simply lie down on the floor and refuse to say anything more.

I'm not going to do any of those things, because I have no desire to do them. But I do have these raw abilities. They involve movements I am

capable of executing, unimpaired in the sense given, the exercise of which is a way of bringing about the accomplishments described, in the circumstances I am in.

Clause (iii) of the analysis involves the *way-of* relation. This is a relation that holds among three abstract objects: a type of movement, a type of circumstance, and an action. The way I use the terms, acts are particulars and actions are types. Actions that agents performs by executing more basic actions, ultimately by executing movements, I call ``accomplishments''. The word suggests something one might be proud of, but I don't mean it in that way. If I were to lean too far over the podium, topple off the stage, and break my arm, my wife might say, ``My, look at all you have accomplished this evening." She would be using the word `accomplishment' in the way I use it in this talk. Accomplishments are things one *does* by doing other things, ultimately, by moving one's body and its parts in various ways, whether those things are intended, or admirable, or not.

Certain types of movements bring about certain kinds of accomplishments in certain kinds of circumstances. A certain movement is a way of *reaching and grabbing this Brussels sprout*, in the circumstance in which this Brussels sprout is located in a certain distance and direction from the agent, no one else is going to grab it, and so on.

It is a fact about the world as a whole, or about certain parts or regions of the world, that executing certain types of movements in certain types of circumstances, are ways of accomplishing certain things. Moving a glass in a certain manner is, near the surface of the earth, a way of getting a drink, in the situation in which there is water in the glass. Out in space, the same movement, in the same circumstance, might not be a way of getting a drink, but a way of splashing water all over one's face. Mailing in a ballot is a way of voting in California, but not in some other places. And so forth. Facts of this

sort, facts relations that hold among types of events, Jon Barwise and I called “constraints” in our book *Situations and Attitudes*.¹⁵

The analysis I am putting forward is not a conditional or dispositional analysis. Certain subjunctive conditionals, such that if I were to execute a certain movement, I would pick up and eat this Brussels sprout, seem to follow from fact that executing that movement is, in my present situation, a way of eating the Brussels sprout. But the analysis I am offering is a categorical analysis. Certain constraints hold, in the actual world; that is, certain relations obtain among types of events or situations. Kissing *involves* touching. Scratching one’s head *involves* bending one’s arm. Breaking an egg in a hot pan and pushing it around for a while with a fork is, in certain circumstances, *a way of* making scrambled eggs. And moving one’s arm and hand and mouth in a certain coordinated way is a way of picking up and eating a Brussels sprout on the podium in front of one. Facts about a person’s muscles and know-how, and the impaired or unimpaired state of their joints and muscles, are facts about the real world. And because of these facts, some people, in some circumstances, *can* do things that they won’t in fact do. That is, in the actual world, I *can* eat this Brussels sprout because I have a certain complex property, in virtue of having unimpaired competences, being in certain circumstances, and facts about the way-of relations that obtain in the world as a whole, and in the part of it I inhabit.

THE CLASSIC CONSEQUENCE ARGUMENT

The consequence argument basically says: You can’t change the laws of nature. You can’t change facts about the past. So if facts about the past and the laws of nature determine that you will perform a certain action, there is nothing else you can do.¹⁶

¹⁵ Cambridge, MA: MIT-Bradford, 1983; reprint with additions, Stanford: CSLI Publications, 1998.

¹⁶ See Peter Van Inwagen, Peter. An Argument for Incompatibilism,” in Watson, op.cit.: 38-57.

The consequence argument can be seen as a development of a more primitive incompatibilist argument in the light of compatibilist criticisms, as John Martin Fischer has pointed out.¹⁷ The more primitive argument goes like this: Suppose it is a law of nature, or as we might say, *naturally necessary*, that if things were a certain way in the past, you will do *A* now. And suppose things were that certain way in the past. Then it is naturally necessary that you will do *A* now, and so you can't refrain from doing so.

The compatibilist reply is that the argument is fallacious, of the form:

$p \ \& \ N(p \rightarrow q)$, therefore $N(q)$.

It is a fallacy to reason from the necessity of a conditional, and the truth of the antecedent, to the necessity of the consequent. It's necessary that if you scored more points than I did, I lost. And you did score more points. That doesn't make it necessary that I lost. I lost for purely contingent reasons, and it was possible for me to win, I simply didn't.

The consequence-arguer replies: Yes, *that* argument was fallacious. But an argument of the following form would *not* be fallacious:

$N(p) \ \& \ N(p \rightarrow q)$, therefore $N(q)$.

One can reason from the necessity of a conditional, *and* the necessity of its antecedent, to the necessity of its consequent. If, for some reason, it was a necessary truth that you scored more points than me, then since it is necessary truth that if you scored more points than me, I lost, it would be a necessary truth that I lost. And, the consequence arguer continues, in the consequence we do have $N(p)$, as well as $N(p \rightarrow q)$, because *p* is a symbol for a statement about the past, and the past *is* necessary: one *cannot* change the past.

¹⁷ In his Cornell dissertation, *Contemporary Approaches to Free Will* (1982) and also in *The Metaphysics of Free Will*, Oxford: Blackwell, 1994. In the latter Fischer develops his own version of the consequence argument, which I criticize in "Can't Can't We All Just be Compatibilists?: A Critical Study of John Martin Fischer's My Way," *Journal of Ethics*, 12, June 2008: 157-166.

Against this, the compatibilist can point out that the argument in question really involves three kinds of favored status, all being called by the name “necessity”. There is natural necessity, that is, the necessity of laws. Then there is the sort of necessity that past events have relative to later times, and finally the sort of necessity that actions one cannot refrain from doing have. So the argument is really of the form:

$$N_{\text{PAST}}(p) \ \& \ N_{\text{LAWS}}(p \rightarrow q), \text{ therefore, } N_{\text{CANNOT-REFRAIN}}(q)$$

But this isn’t obviously correct. It all depends on the relations among these different sorts of favored status that propositions can have --- calling them all ‘necessity’ doesn’t clear this up. The argument really involves what Fischer calls a “transfer principle”, a substantial principle which does not fall out of modal logic. The question is basically this: If a proposition has been made true by events that have already occurred as of time t , then it has a certain favored status at t and all later times. Nothing done at those times can affect the truth of the proposition. If such a proposition together with laws of nature entail the proposition that a person will do something at or after t , does that proposition then *also* have the status that nothing done at those times can affect its truth-value?

It seems clear that the proposition about what has happened before t does not have *exactly* the same status as the proposition that says what the agent will or won’t do at or after t . The former is *already* made true by events that have occurred before t , and the latter doesn’t have *that* status. So *that* status is not transferred. So the transfer principle is at least not trivially true. However, it is also not obviously false. It will depend on exactly how these various statuses that propositions can have relate to one another, and especially how the can-properties work. In particular, if the compatibilist analysis of can-properties is correct, the transfer principle will not hold. The incompatibilist needs to find something amiss with this analysis.

GINET'S ARGUMENT

The best argument that I know of for the transfer principle is given by Ginet in *On Action*. And a version of Ginet's argument can be deployed in terms of the analysis I have given of raw abilities, which is not that different from analyses he considers. So I will now state the argument, and then say what I think is wrong with it. This won't exactly be definitive, since there may be other better ways Ginet can state his argument, and I certainly won't discuss all the points he brings to bear. But it's the best I can do in the time I have this afternoon.¹⁸

I'll repeat my basic analysis. An agent has *the raw ability* at to *A* if:

- (i) There is a type of movement *M*, that the agent has the unimpaired competence to execute;
- (ii) The agent is in circumstances of a certain type *C*;
- (iii) Executing *M* in circumstances of type *C* is a *way of A-ing*.

Ginet's argument is based on what he calls *the principle of the fixity of the given past*. As I am reconstructing his argument, this is a principle about the circumstances that figure in (ii) and (iii).

When we say that the agent is in circumstances of type *C*, what exactly should we have in mind? One alternative is that *C* should contain *all* the facts, past, present, or future. But if we do this, we will be open to a fatalistic argument; that is, an argument that we can't do anything we are not in fact going to do, independent of determinism:

- Suppose, for the sake of argument, that the agent will not *A*, but nevertheless *can A*.
- Then, on the present proposal, part of the circumstances of the agent is that the agent will not *A*.

¹⁸ Which of course does not strictly imply that I could do any better with unlimited time.

- If executing M is a way of bringing it about that P in circumstance Q , then executing P is a way of bringing it about that $P \& Q$ in circumstance Q . (Call this the enrichment of results by circumstances).
- If the agent is in the circumstances in which he will not A , but nevertheless can A , then he can bring it about that he A s and that he doesn't A .
- That's a contradiction, so we must give up the supposition.

One can distinguish, however, between propositions that have already been made true by events up to a given time, and propositions that haven't yet been made true or false by events up to that time. As of 2010, events have made it true that Obama was elected in 2008, but they haven't made it true that he is elected in 2012, nor have they made it true that it's not the case that he is elected in 2012. It seems that the circumstances within which one acts, at a given time, ought to be restricted to things that have already happened. But should it include all of the circumstances, or only part of them? Either restriction will enable us to avoid the fatalist argument.

Ginet says that circumstances ought to include *all* the facts fixed by what has happened up to the very moment of action. That is, if Q is a complex proposition that has incorporates everything that has been made true by events that have occurred up to a time t , then Q is the circumstance in which any act at t occurs. So, Ginet says, that should be the circumstance we are interested in, in clause (ii) and (iii); executing M should be a way of A -ing in the totality of circumstances fixed by the (actual or given) past.

With this principle and determinism, we get an argument very much like the fatalist argument. Let C^t be the totality of circumstances fixed by events up to t .

- Suppose, for the sake of argument,
 - that the agent does not A at t ;

- the agent *can A* at *t*;
- determinism is true;
- Then, *C^t & the agent As at t*, is a counter-instance to the laws of nature. .
- Then by the principle of the enrichment of results by circumstances, the agent can bring it about that *C^t & the agent As at t*.
- Then the agent can bring about a counter-instance to the laws of nature;
- But no one can do that, so either the agent *As*, or it's not the case that the agent can *A*, or determinism isn't true.

Suppose it is a (non-basic) law of nature that no one who really dislikes Brussels sprouts, and has no reason whatsoever to eat a Brussels sprout, ever eats one. Suppose I really dislike Brussels sprouts and have no reason to eat one. Still, the compatibilist claims, I can eat this Brussels sprout because it is here on the podium and I have the competence to move in the right way to eat it.

Given Ginet's principle, that would also mean that executing the movement in question is a way of bringing it about that I eat a Brussels sprout in the circumstance that there is a Brussels sprout on the podium in front of me and I really dislike Brussels sprouts and have no reason to eat one, for those are all part of the total circumstances. But then it seems I can bring it about that I really hate Brussels sprouts and have no reason to eat one *and* that I eat one. But then I can bring about a counterexample to the law of nature we imagined. But we cannot bring about counterexamples to the laws of nature.

With regard to both the fatalist and incompatibilist arguments, the compatibilist might consider giving up the principle of the enrichment of results by circumstances. Suppose one can bring it about that *P* in circumstance *Q*. Does it really follow that one can bring it about that *P & Q* in circumstance *Q*? If I eat the Brussels sprout, will I have brought it about that I have eaten the Brussels sprout *and* Caesar conquered Gaul? If we can deny

that, we can stop both arguments. In the second one, we can say that although I can bring it about that I eat the Brussels sprout *in the circumstance* in which a really dislike them &c., it doesn't follow that I can bring it about *that* I eat the Brussels sprout *and* I really dislike it &c. Then we can deny the further step, that I can bring about an exception to the law of nature.

The principle in question follows naturally from what I call an incremental conception of the results of action. Action is by its nature *completing* conditions that suffice to make things happen. You don't have to bring about all the elements in the sufficient condition, for getting the thing to happen. You just have to make, so to speak, the final contribution. If we had a thousand dollars in our bank account, and you spent \$950 last week, and then I wrote a check for \$100 this week, I have overdrawn the account. I didn't bring it about that we only had \$50 left, but I brought it about that we only had \$50 left and I wrote a check for \$100. I certainly don't deserve all the blame for our predicament, but I am the one who overdrew the account.

So, I don't want to deny the principle of the enrichment of results. I go along with Ginet, in thinking that the circumstances in parts (ii) and (iii) of our analysis ought to be restricted to past circumstances, so fatalism isn't a problem. But I don't think we should include *all* of the past circumstances; that is, I disagree with the principle of the fixity of the given past.¹⁹

¹⁹ In his ingenious unpublished paper "Freedom and the Fixity of the Past," Wes Holliday develops an argument in favor of incompatibilism based on the plausible idea that no one can do anything that no one does in any possible world. I think, although I am by no means certain, that the considerations I provide in response to Ginet's account can form the basis of a plausible reply to Holliday's argument, but I cannot pursue that here.

THE ENRICHMENT OF CIRCUMSTANCES

Suppose that C and D are parts of the total circumstances C^t , that is, the total circumstances made true by events up until t . And suppose,

- (i) Executing M is a way of bringing it about that P in circumstance C .

Does it follow that

- (ii) Executing M is a way of bringing it about that P in circumstance C & D ?

For example, if executing M in the circumstance in which there is a Brussels sprout before me on the podium is a way of eating the Brussel sprout, does it follow that executing M in the circumstance in which there is a Brussels sprout before me on the podium *and* I really don't want to eat a Brussel sprout, is also a way of eating a Brussel Sprout?

It seems that we can derive Ginet's principle from this one, which I'll call "the enrichment of circumstances," by simply taking the additional circumstance D to be everything else in addition to C that is fixed by the past. An this principle allows us to focus on what seems to be the real issue, which is the question whether I can, say, eat this Brussel sprout given that, as we may suppose, my preference structure and the laws of nature settle that I am not going to eat it. I think we should deny the principle of enrichment of circumstances.

When executing M is a way of bringing it about that P in circumstances C , the combined circumstance of C and the execution of M is naturally sufficient for P . That is, the way things work, whether deterministic or not, is

such that this combination will suffice for bringing it about that P .²⁰ With respect to this sufficient condition for P , C is what Mackie calls an INUS condition; an *Insufficient* but *Necessary* part of an *Unnecessary* but *Sufficient* condition for P . That is, given the way the world works, M and C form a sufficient condition for P . C gets at what is necessary to *close the gap* between executing M and bringing it about that P . Suppose I move my arm, hand and mouth in the requisite manner. What else is required of the world, given the way things work, for this to be a case of bringing it about that I eat the Brussel sprout? Basically, that Brussel sprout has to be a certain distance and direction from me. And it can't be glued or nailed to the podium. And there can't be some errant cosmic ray that will turn it to dust on the way to my mouth. C is a complex condition, but in spite of its complexity, I am good at detecting whether it holds. The core parts of C can be checked visually, and the rest, like the absence of cosmic rays and the absence of glue, are pretty likely.

Now if this, or something like this, is the right way to look at the conditions involved in instances of the way-of relation, then the principle of the enrichment of circumstances should be rejected. The status of being an INUS-condition, relative to a movement and the way the world works, does not persist through enrichment. Suppose C is the condition that makes a certain movement a way for me to eat this Brussels sprout. Then how about the complex condition consisting of C *and* the fact that Caesar conquered Gaul. That complex condition is not an INUS condition, because it is not necessary. This movement plus the circumstance that the unglued Brussels sprout is where it is suffices to bring about my eating the Brussels sprout, whether Caesar conquered Gaul or not. That is, the property of being a condition that is necessary for some result does not *persist* through enrichment.

²⁰ "The way things work" includes not just laws of nature, but also conventions and other constraints that are relevant to what Alvin Goldman calls "act-generation". See his *Theory of Human Action*, Englewood Cliffs, New Jersey: Prentice Hall, 1971.

Thus the compatibilist can say that my executing a certain movement is a way of eating this Brussels sprout in the condition that it is where it is, isn't glued down, etc., without having to claim that it is a way of eating this Brussels sprout in the circumstance that combines this with the fact that I really desire to not eat the Brussels sprout, and have no reason to. When I think that I can eat this Brussels sprout, my thought is true, because there is a circumstance I am in, and there is a movement I have an unimpaired competence to make, such that executing the movement is a way of eating this Brussels sprout in the circumstance. The fact that I don't want to eat the Brussels sprout is surely relevant to the issue of whether I *will* eat it, but it is no more relevant to the issue of whether I *can* eat it, than the fact that Caesar conquered Gaul.

DELIBERATION AND THE FIXITY OF THE GIVEN PAST

It may be that although there are no decisive logical reasons for accepting the principle of the fixity of the given past, there are compelling phenomenological reasons for doing so. Ginet says this:

My impression at each moment is that *I at the moment*, and nothing prior to the moment, determine which of several open alternatives is the next sort of bodily exertion I voluntarily make (90).

The word 'nothing' does the incompatibilist's work here. Suppose I have the impression that I can do *A* and I can also refrain from doing *A*, and in fact I refrain. According to my analysis, my impression is true if I have an unimpaired competence to execute a movement that is a way of doing *A*, in some circumstance I am in, and also an unimpaired competence to do execute a movement that is a way of refraining from doing *A* in some circumstance I am in. But what counts as circumstances I am in? If only the totality of *everything* fixed up until the time of action counts --- if *nothing* can be left out --

- then it seems that, given determinism, my circumstances *plus* the laws of nature determine that I will refrain. But then, if I can do *A*, I can do *A* in the circumstance being in that state that together with wider circumstances and the laws of nature determines that I will refrain from doing *A*. But then I can violate the laws of nature. But I can't do that.

So our impression when we are free, expressed with, "I can to *A*, but I can also refrain from doing *A*," if correct, rules out determinism. Given determinism, unless we leave *something* out of the circumstances, the circumstances plus the laws of nature settle what we are going to do. And, at least according to Ginet, *nothing* can be left out.

It seems to me, however, that the condition Ginet sets off with commas, "and nothing prior to the moment," is phenomenologically incorrect. That is, this does not seem to be a part, or, as Ginet puts it, an "essential ingredient" of, the impression of freedom.

Suppose I'm at the grocery store. My wife has asked me to pick up a carton of lite (1% fat) milk. Now I stand before the milk cooler. In a moment I will select a carton of skim milk or a carton of lite milk and put it in my cart. I can't remember which she said to get. I have the impression that am free to get either one. I might say or think to myself, "I can get the lite, and I can get the skim." However, I know that she told me to get one or the other, and that I wrote it down, and put the note in my pocket. I know that I will buy whichever one I find that I was told to buy, when I look at the note. My belief may not be connected with determinism. I just know that I will decide in accord with what is already written on the note, because I want to please my wife, and when I read what is on the note, I will come to believe either the proposition that I should get the skim milk and not the lite milk, or the proposition that I should get the lite milk and not the skim milk. So it seems to me that my impression definitely *does* involve, as an essential ingredient, that it is I who determines which bodily exertion I voluntarily make. But it

does *not* involve the restriction that nothing prior to the moment determines this. What is written on the note is going to determine what I do.

Now suppose that I do remember what she said; she said, "Get the lite milk, not the skim milk". As I walk to the milk cooler, I still have the feeling of freedom; it is up to me whether I get the lite milk or the skim milk. If I were in a different mood, I might deliberate among my options. Should I annoy my wife? Would it be worth enduring her displeasure, for one afternoon, in order to foster in her the conviction that it was pointless to give me errands to do? Perhaps on another day, in another mood, such a scheme might tempt me. But not today. I am in a good mood, and don't want to annoy anyone, least of all my wife. Everything is in place for me to go the cooler, pick out the lite milk, and continue with my day. I have the relevant desires and beliefs to motivate that act; I've had them since I left home; there are no serious competing motivations.

Return to the earlier scenario, where I have forgotten what she said, but know I have a note in my pocket that I can consult. Suppose it occurs to me that it may be determined by the laws of nature and the state of the world at this time, before I look at the note, that I will do whichever thing is written on the note; that is, if "skim" is on the note, it is determined that I will get the skim, and if "lite" is written on the note, it is determined that I will get the lite. There might be some not-very-basic law of nature, that when a human being like me is in the mental states I express by saying:

I want to please my wife

I know that she wants me to get the kind of milk the name of which written the note.

I can't think of any reason not to get whatever kind of milk the name of which is written on the note

and, in addition,

"Lite milk" is written on the note, and stands for lite milk

that human being will decide to get the lite milk? If there were such a law, would that mean that *it wasn't up to me* to get the milk? If I contemplate this, will I come cease to have the feeling that getting the other kind of milk is something I *can* do, that it is *open to me*? Or will I continue to have the feeling, but think that it is an illusion? These seem to be quite misguided conclusions. Shouldn't I simply think it is something I can do, but I'm not going to do, because I don't want to? I want to get the kind of milk my wife asked me to get, and don't want to get the kind she didn't ask me to get, but I can get either of them; it's up to me.

So let's assume that this is my state of mind. I have the impression of freedom; I think it is up to me which carton of milk I get; it is open to me to get the lite or the skim; I have, and know that I have, good reasons to get the lite; I don't have any reason, no urge, no inclination whatsoever to get the skim. Moreover, I think that it is a law of nature that, *ceteris paribus*, people who have every reason to do *A*, and no reason to refrain from doing *A*, no aversion to doing *A*, and are not prevented in any way from doing *A*, will do *A*.

The "ceteris paribus" basically means:

I think there is some condition *X*, that I can't spell out, involving a number of negative conditions to the effect that something untoward isn't going to happen in my brain, or in the milk case, or in the city or state or universe as a whole, that will interfere with the orderly progression of things, and that it is a law that people who meet condition *X* in addition to the other conditions will do *A*, and I believe that whatever *X* may be, I meet it.

Let's hold everything constant except what my wife said. Just assume that my memory and dependability are such that that her statement is the only variable worth considering. Given that my wife said, "Get lite," I will get lite; if I get skim, that would only be because she said, "Get skim". Her saying "Get lite" happened in the past. I certainly don't believe that I can do anything about it now. Nothing I do now, standing in front of the cooler, can

change the fact that half an hour ago she said, "Get lite". And yet, I assume, given that I am a person of type X, and have the desire to please my wife, and so forth and so on, the fact that she said "Get lite" will cause me to get lite. It is a classic INUS condition: an insufficient but necessary part of an unnecessary but sufficient condition. Assume the laws of nature. Add the stuff about my desires and memory and factor X. Given all of that, my wife's saying "Get lite," although not by itself sufficient, is the last ingredient in what adds up to a sufficient condition, for me to choose the Lite.

It seems I would think of things as follows. If I choose the skim rather than the lite, it will annoy my wife. I don't want to annoy my wife. If I wanted to annoy my wife, that's what I would do. Since I don't want to annoy my wife, that's not what I am going to do. It would be crazy to pick the skim and not the lite, unless I wanted to annoy my wife. My not being crazy is part of factor X. So it is *not* open to me to bring about the conjunction of picking the skim and having just the desires and preferences that I do, because the only motivation I would have for picking the skim and not the lite is that I would have different preferences than I do. If I pick the skim, given that my memory is intact, and I am not crazy, that will be because I want to annoy my wife.

So, I suggest, a close study of the phenomenology of the impression of freedom does not support the idea the principle of the fixity of the past is a true principle about "can", "open to" "have the power that" and the like, in the very sense in which they occur in deliberation

So now I have looked at the note, seen that it says "lite", and am ready to reach into the cooler and grab a carton of lite. It's perfectly clear to me what I am going to do. But I ask myself, *could* I get the skim? It seems I would answer "Yes, I can get the skim, although I won't". Why can I get it? Because it's right there in front of me. The store isn't out of skim milk; it's not too high on the shelf, or too far back, for me to reach. All I would have to do to get the

skim is to move my hand and arm in a certain way, and I certainly have the competence to do that.

REVISIONISM

I taught a seminar about all of this last quarter at Stanford.²¹ Among other things, we used the book by *Four Views on Free Will*, by John Martin Fischer, Derek Perebom, Robert Kane, and the brilliant young philosopher from the Bay Areas's very own University of San Francisco by way of Stanford, Manuel Vargas. Vargas defends a view he calls "Revisionism".²² In addition, Vargas attended the seminar. I have taught the subject of freedom and determinism in some course or other almost every year for the past forty years or so. I always argue for compatibilism, and the result is almost always that more students are incompatibilists by the end of my efforts than were at the beginning. The exception was this last seminar. By the end of the seminar, most students were revisionists. This wasn't because Vargas dominated the seminar and drummed it into their skulls. He was very well-mannered and was very deferential to my views. One possible explanation of this is that his view is quite plausible. Another is that the students thought that Revisionism captured everything that was plausible about my view, but in a kinder, gentler way. So I want to end by saying why I don't just describe my view as revisionism.

Vargas begins his essay with some examples of concepts that we have had to revise in the face of changes in knowledge and culture. The first one is water, which I mentioned at the beginning. Once philosophers thought that water was a basic element. Now we all accept that it is a compound. Just as we have had to revise our conception of water, and our conception of marriage, and our conception of, say, what a wife is, we need to revise our conception of freedom. We know longer think that water is a basic element.

²¹ Winter, 2010.

²² Oxford: Blackwell Publishing, 2007.

We no longer think that wives are chattel. The current conception of freedom, in the minds of most ordinary users of the concept, is closest to the one the incompatibilists have, Vargas thinks. It should be revised in favor of the compatibilist conception, which is more rational and practical and useful, given what we know about the way the world works.

Grant for the sake of argument that some incompatibilist's conception of freedom reflects what most people think about it. Then I agree with Vargas that they should revise their opinion. But I think the case is more like water, and less like marriage or being a spouse, than he seems to. Freedom and the can-properties are not social constructions, but properties humans, and other animals, learned to recognize early in the process of the evolution of intelligence. Water is the stuff that plays a certain role in human life, and did so long before there was a word 'water', or any of the various human institutions that now deal with water: purification plants, dams, canals that enable you Southern Californians to steal water from us Northern Californians to fill up your swimming pools, and the like. The word 'water' stands for water because it is the stuff that plays that central role in our lives, not because it is the stuff that best fits some set of opinions that deep thinkers or the general public associates with 'water'. The idea that water was a basic element no doubt reflected the view that humans have such a central role in the scheme of things, that any substance that plays such a central role in their lives, and is indeed a necessary condition of human life, must be basic to the wider scheme of things. But that view is just wrong; false; in error.

I think things are quite parallel with the can-properties, the properties the predicates we create with 'can' and verb phrases stand for, that we think about and deliberate about in deciding what to do, and predicting and explaining and evaluating what others do. These properties played a large role in human life before anyone had any conception of them, or word for them. The incompatibilist theory of them is simply wrong; even if everyone thinks it is the right conception of them, it is not. It is, I think, largely a result

of confusing the concept of free will, a philosophical/religious construct, with the natural property of freedom.

I do not think it is entirely clear from Vargas's essay whether he advocates revisionism in the sense I do --- people who have the wrong opinion about what freedom is should revise their opinion --- or in a sense with which I don't agree. On such a view, this wrong conception determines the present referent of 'freedom' and the various 'can' predicates; we should revise our conception of freedom, and thereby change what this useful word stands for. I am a revisionist in the first sense, but not the second.

That said, last Fall's seminar was the most successful I have taught, in that most of the students ended up being revisionists rather than outright incompatibilists. And if this lecture has a similar result, I'll be most happy.²³

²³ I became a compatibilist in 1968, due to conversations with John Taurek. Since that time I have had numerous illuminating conversations with many students and colleagues, including Michael Bratman, John Martin Fischer, Wes Holliday and Manuel Vargas among many others.