Determinism and the Case of the Contrarian

Jesse Spafford

Pomona College, Class of 2012

Can free will be saved from the threat of determinism? If one is willing to embrace a very modest notion of free will, certainly. However if one subscribes to a strong conception of free will, one that demands that a person could have been able to act otherwise than she did, then the picture looks much bleaker. For such a strong view requires that one's actions be contingent while the determinist thesis argues that one's actions are, in fact, necessary. In "The Incompatibility of Free Will and Determinism," Peter van Inwagen states the determinist thesis as follows:

(a) For every instant of time, there is a proposition that expresses the state of the world at that instant.

(b) If A and B are any propositions that express the state of the world at some instants, then the conjunction of A with the laws of physics entails B. [1]

The problem such a view poses for a strong account of free will is that it appears to be correct. All non-living objects certainly obey physical laws and only transition from one state into another in accordance with those laws. A ball placed on a flat plane will not *spontaneously* begin to roll and swerve about. Rather, such behavior would only occur if the appropriate forces were applied to the ball in order to put it into motion. Further, there seems to be no reason to draw a distinction between what is living and what is non-living as far as physical laws are concerned, for, while living organisms may certainly be more complex than non-living objects, they are both comprised of the same basic building blocks of matter. While the way in which the atoms in the brain respond to a stimulus or external force is far more complex than how a ball responds to a stimulus like a push, their behavior—and, thus, our behavior—would seem to be the necessary effect entailed by the initial stimulus in combination with physical laws.

However, while this deterministic understanding of reality is compelling, in this paper I will argue that, unless we are willing to reject premises even more plausible than the determinist thesis, the latter will culminate in a contradiction and should therefore be

---

[1] van Inwagen, Peter, 1975, "The Incompatibility of Free Will and Determinism", *Philosophical Studies*, 25: 185–99. p. 186

rejected. While showing that the thesis is false in this way will not serve as a positive proof that we *have* free will, it will undermine one of the primary reasons for thinking that we do not, thereby preserving the possibility that we have free will.

The contradiction inherent to the determinist thesis manifests itself in what I will call the Case of the Contrarian. In this case, scientists living in a world which we will assume is deterministic have managed to build an incredibly powerful Location And Physical Law Analyzing Computer Extraordinaire (LAPLACE) that is able to simulate physical laws such that, when the scientists enter the state of the world at a given time t, it is able to simulate how the atoms at t will interact and, thus, where and in what state they will be in at a later time t'. In this way, if the initial conditions of the Big Bang were entered into the computer, it would be able to accurately simulate the entirety of human history up through the present moment and indefinitely into the future.

The scientists responsible for creating LAPLACE are ecstatic. They have finally built a device that can see into the future with perfect accuracy by simply simulating the position and movement of atoms. Now, the only one thing left to do is to try out the machine. After entering the initial conditions of the Big Bang—the scientists had spent years to deducing them from current conditions—they put LAPLACE into motion. The computer begins churning away, the scientists pausing it occasionally to see if it is simulating historical events accurately—and indeed it is. Pleased with their work, the scientists take a leisurely lunch break and return to the lab to find that LAPLACE has simulated time all the way up through the present and into the future.[2]

One scientist, Susan, pauses the computer. Unable to contain her curiosity, she decides to search for where the cluster of atoms that make up her body will be in three hours. The computer quietly whirs for a few moments and then pulls up results indicating that, in three hours, her body is sitting at a coffee shop playing checkers with a man she does not recognize. Susan, however, is a contrarian and decides that she will do no such thing. Thus, after roughly two hours have elapsed, she drives the thirty minutes home, sits in her living room, and watches television for the rest of the day, thus invalidating LAPLACE's prediction.

This case is very strange in that its premises, including the one stating that the world is deterministic, all seem reasonable and yet they clearly lead to a logical contradiction: LAPLACE is a computer that accurately predicts the future and yet it cannot accurately predict the future. Thus, one of the premises, while appearing to be valid, must be rejected as false in order to avoid this contradiction. To investigate which of the premises is the fallacious one, I will begin by laying out the premises of the case

---

[2] Of course, the scientists have no way of verifying that LAPLACE's predictions for the future are correct, given that the events have yet to have been actualized. However, given that the universe is governed by physical laws and the computer can accurately simulate those laws (premise 3, below), it follows that the computer's predictions will be accurate. Further, the scientists seem to have some justification for believing the predictions to be correct given LAPLACE's remarkable ability to simulate the past with total accuracy.

and, through process of elimination, will reject the determinist thesis as being false. The basic premises of the Case of the Contrarian are as follows:

1. It is possible for the computer to be physically constructed such that it is operable.

2. It is possible to ascertain the initial conditions of the universe and enter them into the computer.

3. Given initial conditions, future conditions can be *accurately* predicted using physical laws (i.e., determinism obtains)

4. The computer has the capability to simulate these physical laws and, thus, predict future conditions.

5. The computer can report the future conditions to the scientists such that they learn the full extent of the computer's predictions.

6. Given Premises 1-5, the future conditions are accurate such that if the computer says X will occur, then X occurs and if the computer says a person will Y, then she Ys.

7. Susan is a contrarian such that, if she is told she will Y, then she does not Y (~Y).

The contradiction arises from Premises 6 and 7. It is not possible for Susan, if told she will Y, to both Y, as Premise 6 entails, and ~Y, as Premise 7 entails. As Premise 6 follows from the first five premises, this contradiction means that one of the seven premises must be false. The question is, which one?

Let us begin by questioning the validity of Premise 7, as it appears to be the premise that generates the contradiction in the Case of the Contrarian. An easy way out of the logical inconsistency presented above would be to simply deny that Susan could have not gone and played checkers at the coffee shop. However, such a proposition—that, when told one is going to Y, one necessarily Ys—seems to fly in the face of reality as we understand it. While perhaps there are some actions that, when predicted, one will do out of necessity (e.g., continue breathing, eating, etc.) LAPLACE should also be able to predict all the trivial things one does in life such a get coffee or play checkers. It seems inconceivable that a person, upon receiving a computer printout informing her that she will play checkers, has no choice but to play checkers.

Further, there seems to be nothing problematic about the stipulation that the scientist in question is a contrarian such that she will ~Y. Such a person would be analogous to a machine that, upon detecting Y, was programmed to ~Y.[3] There certainly

---

[3] I am indebted to Peter Kung for this suggestion.

seems to be nothing problematic about such a machine, and the same seems to hold true for the contrarian.

Also, to avoid charges of question-begging, I should point out that premise 7 does not presuppose any sort of free will or non-deterministic metaphysics. Even if our thoughts are determined by our environment, it does not seem unreasonable that there could be an individual whose actions are determined in such a way that, when told what she will do, will always do the opposite. Such a contrarian does not necessarily have a choice in the matter nor does she need to have had the ability to do otherwise (i.e., to act in the way in which she was told that she was going to act). Given this, there does not seem to be a good reason why a contrarian could not exist who, when told she will Y, will ~Y. Thus, Premise 7 does appear to be valid despite the contradictions it raises.

If one is willing to grant that Premise 7 holds, Premise 6 must be false as LAPLACE would predict that, "the computer tells Susan that she will Y and she Y's," but, in reality, Susan would be told she would Y and would ~Y. However, we cannot simply discard Premise 6 and declare ourselves finished, as it is the logical conclusion of the first five premises. Rather, one of these first five premises must be incorrect and thus lead to the contradiction between Premises 6 and 7.

Let us begin by considering Premises 1, 4, and 5. I group these premises together because they all pertain to the basic feasibility of constructing and operating a computer capable of predicting determinist laws.[4] Admittedly, it is certainly possible that such a computer could not be built. It might be too large, require more computing power than is possible, or run too slowly to ever catch up to the present and thus predict the future. While these are all legitimate possibilities, I will stipulate away such contingencies on the grounds that they are *in fact* nothing but contingencies: whether or not they are false seems to be merely coincidental to the contradiction between Premises 6 and 7, and, thus, we can legitimately assume them to be true and search for the source of the contradiction elsewhere. To see why I believe this to be the case, consider the following example:

Einsteinian mechanics holds that it is impossible to accelerate an object that was initially at rest to the speed of light.[5] The reason for this is because, as an object's velocity increases, its mass appears to increase as well, although this effect only becomes noticeable when the object is moving very quickly. In fact, as an object approaches the

---

[4] It might be argued that Premise 5 is not a problem with the computer but, rather, a problem of being able to interpret the computer's results since it would be calculating the position of swarms of atoms rather than the macro objects with which we are familiar. However, these macro objects (e.g. coffee, people, checkers, etc.) are merely specific arrangements of these atoms. Thus, LAPLACE would have to be programmed to recognize these patterns and report them to the scientists using their macro monikers. The feasibility of such programming seems to be a question of a similar kind to Premises 1 and 4 and thus I will address the three together.

[5] *Encyclopædia Britannica Online*, s. v. "relativity," accessed May 04, 2012, http://www.britannica.com/EBchecked/topic/496904/relativity.

speed of light, its mass increases so rapidly that it approaches infinity. Thus objects at rest cannot be accelerated to the speed of light because the more massive an object becomes, the more energy is required to accelerate it (consider the amount of fuel needed to accelerate a truck as opposed to a motorcycle). Consequently, as an object approaches the speed of light and its mass approaches infinity, the amount of energy required to accelerate that object grows concomitantly such that, to achieve the speed of light, an infinite amount of energy would be required. As there is only a finite amount of energy in the universe, such a task is physically impossible.

However, imagine that this feature of the universe was unknown to the scientific community and the debate over whether or not it was possible to reach the speed of light was still raging. To try to resolve the dispute, a team of engineers builds a magnetic train on a circular track and attempts to accelerate it to the speed of light. The engineers, of course, don't even come close. However, the reason that they do not get the train to reach the speed of light is not because they lack the energy necessary to accelerate it to such a great velocity; they are unable to get the train up to a velocity close enough to the speed of light for there to be any noticeable increase it its mass. Rather, because of the materials available to the scientists, the engine powering the train overheats whenever the train runs for more than a few minutes, regardless of what material they use to build the engine or what type of coolant they use. The engineers conclude that if the engine of the magnetic train overheats, then any sort of engine will overheat, and all attempts to accelerate an object to the speed of light are abandoned.

So what does this scenario tell us about the Case of the Contrarian? In the discussion regarding whether or not an object can be accelerated to the speed of light, there is the tacit premise that a physical mechanism exists that could do the accelerating without breaking down, a premise that, in the above scenario, turns out to be false. However, in this case there is no physical *law* that precludes such a mechanism from existing. That no such mechanism exists seems to be a contingent truth rather than a necessary one: there is a distinct physical explanation as to why objects cannot be accelerated to the speed of light, to which the problem of engines overheating is merely coincidental. This distinction can be described in terms of explanation: the reason that an object cannot be accelerated to the speed of light is *because* the mass of the object begins to approach infinity and thereby requires infinite energy to continue accelerating it. Thus, in the counterfactual world in which the mass of the object *did not* approach infinity, the object *could* be accelerated to the speed of light. This stands in contrast to the matter of the engine, for, even if the necessary materials for building an engine that did not overheat *did* exist, objects still *could not* be accelerated to the speed of light.

Thus, what I suggest is, with respect to the Case of the Contrarian, while it might be true that Premises 1, 4, and 5 could be false as described above, the reasons why they might be false are contingent to the matter of determinism. Even if it turns out that there could never be enough computing power mustered to run a simulation with a degree of

complexity even close to that of the one described in the Case of the Contrarian, this would be merely a coincidental alignment of a physical truth with the metaphysical question at hand. For, if the two were necessarily related, that would mean that the reason that the world is deterministic is *because* of this lack of computing power or, conversely, that our computing power is limited *because* the universe is deterministic. Such a connection seems implausible: rather, I posit that any limitations to the scientists' computing capabilities are the result of the materials available to them, the strength of electrical currents, etc., and not some necessary relationship between computing power and the laws of cause and effect.

The point of this analysis is to show that it is theoretically legitimate to imagine a world where LAPLACE could be built that does not beg the question regarding the matter of determinism. If there were some necessary connection between computing power and determinism, then assuming the truth of Premises 1, 4, and 5 would entail the falsity of Premise 3. In such a case, the *reductio* I am trying to establish could easily be dismissed by the determinist as being as trivial as an argument that purports to disprove Einsteinian mechanics by assuming the falsity of Einsteinian mechanics as a premise. However, if I am right in thinking that it is not obvious that such a necessary connection exists, then we can presuppose the truth of Premises 1, 4, and 5 and still be surprised if Premise 3 turns out to be false, as one of the other premises might equally well have been responsible for the contradiction between Premises 6 and 7.

Let us then turn our attention to Premise 2. The objection to this premise would go something as follows: there is no way to know the initial conditions of the universe, as humans simply don't have the tools to predict these origins with any accuracy and likely never will. There are two responses to this objection. First, this limitation, too, does not seem causal to the contradiction raised between Premises 6 and 7. Second, while it may not be possible to know for certain the initial conditions of the universe, we can imagine a world in which the scientists building the computer could come close using available tools and *accurately* guess that which they cannot determine. Such a guess would be beyond lucky, but not impossible. Thus, for the Case of the Contrarian, we can dismiss objections to Premise 2 by stipulating that, in this one case, the scientists managed to make that incredibly lucky guess and therefore obtained the initial conditions of the universe. While people in this hypothetical world may doubt the accuracy of that guess, their skepticism would be irrelevant because Susan, the contrarian, will act counter to whatever the computer says she will do regardless of whether or not she believes the computer's predictions to be correct. While Susan may not prove anything to those who believe the initial conditions were incorrect, we external observers know that, in this case, the initial inputs into LAPLACE were, in fact, correct, and, thus, her ability to act contrary to what the computer says is highly significant whether the people in that world recognize it or not.

So where does this leave us? If Premises 1, 2, 4, 5, and 7 are all true (with Premise 6 being the logical conclusion of Premises 1-5), then the conclusion that one must draw is that Premise 3—the premise that posits the determinist assumption that, given initial conditions, future conditions can be accurately predicted using physical laws—must be rejected in order to avoid the logical contradiction between Premises 6 and 7. Why determinism fails is unclear. It is certainly not justified to claim that this case is an example of free will as it is equally possible that randomness at the quantum level undermines the strict determinist understanding of cause and effect rather than human agency. However, given that the determinist thesis is one of the primary tools in the arsenal of those who would deny free will, the revelation that determinism is false due to an entailed contradiction would seem to shift the debate in favor of those who believe that we can do otherwise.

Perhaps free will can be saved after all.