20

A Reductionist Account of Personal Identity[1]

Fauve Lybaert

Descartes, René. *Meditations on First Philosophy*. New York: Classic Books America, 2009.
Locke, John. *An Essay concerning Human Understanding*. Indianapolis: Hackett, 1996.
Nagel, Thomas. *The View from Nowhere*. Oxford: Oxford University Press, 1986.
Parfit, Derek. "Experiences, Subjects and Conceptual Schemes," *Philosophical Topics* 26, 1/2 (1999): 217–70.
____. "Is Personal Identity What Matters?" The Ammonius Foundation. http://www.ammonius.org/assets/pdfs/ammoniusfinal.pdf (accessed December 31, 2007).
____. *Reasons and Persons*. Oxford: Oxford University Press, 1984.
____. "The Unimportance of Identity," in *Identity*, edited by H. Harris, 13–46. Oxford: Oxford University Press, 1995.
Quine, W. V. "Identity and Individuation." *The Journal of Philosophy* 69 (1972): 488–97.
Shoemaker, Sydney. "Persons and Their Pasts." *American Philosophical Quarterly* 7 (1970): 269–85.
Williams, Bernard. *Problems of the Self*. Cambridge, UK: Cambridge University Press, 1973.
Wittgenstein, Ludwig. *Zettel*. Oxford: Blackwell, 1967.

[1] The exposition of this account draws heavily on the work of Derek Parfit. The exposition of the different kinds of reductionism is in large part inspired by Parfit's "Experiences, Subjects and Conceptual Schemes" as well as his "Is Personal Identity What Matters?" in which he slightly revises the argument which he makes about personal identity in *Reasons and Persons*. The formalized argument at the end of this chapter is an abbreviated version of the argument which Parfit develops in his *Reasons and Persons*. Both the commentary and the formalized argument have benefited from the comments of Derek Parfit, Cheryl Chen, Filip Buekens, Lorenz Demey, and Roger Vergauwen.

Just the Arguments: 100 of the Most Important Arguments in Western Philosophy,
First Edition. Edited by Michael Bruce and Steven Barbone.

For ages, philosophers have argued over the nature of persons and what is involved in the numerical identity of persons over time. To understand the concept of numerical identity, consider this. The two chairs at my kitchen table, which look exactly alike and are made of the same material, may be qualitatively identical, but they are not numerically identical. Contrast this with the one chair in my room. If someone paints that chair while my eyes are closed, then the chair I see when I open my eyes will be qualitatively different from but numerically the same as the chair I saw before.

Apply this to persons. When a relative tells you that you have changed over the years, he recognizes that you are still numerically the same person. He does not think that you have passed away. But he sees that you are qualitatively a bit different now.

There is more debate over whether someone is still numerically the same person when complete loss of memory and radical change of character occur. Philosophers disagree over whether the resulting person is only qualitatively different or also numerically different than the person before having a brain hemorrhage. Philosophers, such as Derek Parfit, who hold that we are only the same as long as there is psychological continuity, say that we would in such a case be confronted with a numerically different entity. Philosophers, such as Bernard Williams, who state that someone stays the same as long as there is bodily continuity, claim the opposite.

How do we decide what determines the numerical identity of someone? We will first have to agree on how the concept "person" gets its meaning. John Locke (*An Essay*, 148 II.xxvii.26) stated that the concept 'person' is a forensic concept. "Forensic" is often equivocated with "legal", but its meaning stretches further than this. The term is derived from the Latin term '*forum*' and means "public". Locke refers to "person" as a public concept because he takes its meaning to be determined by how we use it – or, to be more precise, by how we ought to use it if we want our speaking to be in accordance with our common beliefs, attitudes, and practices. The meaning of the word 'person' in a legal context is one instance of this. It has, for example, been held that, in this context, someone cannot be found guilty of committing a crime unless he remembers committing it. One idea behind this is that it only makes sense to penalize someone for doing something if he can take responsibility for doing this. Remembering what you did is supposed to be a precondition for the latter.

However, not all philosophers agree on whether the meaning of the concept "person" is determined by our common use of it. Derek Parfit, for instance, contests this assumption. He warns that our use of this term may be wrongheaded and holds that philosophers are in a position to assess this. They can unveil inconsistencies in our use of this concept, examine whether there is a real entity in the world to which it refers, as well as determine

whether this concept names what matters when we are concerned about our survival – as we usually think it does.[2]

This being said, philosophers will mostly start their examination of what the concept "person" refers to with an assessment of how we commonly use this concept. They will either describe our use of this concept as precisely as possible and let this description function as a determination of the meaning of this term, or they will explain why our application of this concept is not entirely accurate.

This has led to two main philosophical approaches to the questions of what persons are and what makes a person maintain her numerical identity over time: the reductionist and the nonreductionist approach.

There are different versions of reductionism. Constitutive reductionism[3] is likely to be the most defensible version of reductionism with regard to persons. Constitutive reductionists admit that persons exist but argue that they are fully constituted by their physical and/or psychological continuity, and nothing over and above these continuities.

To say that persons are fully constituted by their physical and/or psychological continuity is not to say that persons are nothing but this continuity. According to Sydney Shoemaker, the case is analogous to the relationship between a statue and the lump of clay of which it is made. The statue is constituted by the clay and has no separate existence apart from the clay. Yet it is not the same as the lump of clay. For, if this lump loses its shape, it will still be there, but the statue will not be.[4]

Constitutive reductionists are metaphysical reductionists, not conceptual reductionists.[5] They claim that persons are not separately existing entities over and above their physical and/or psychological continuity, even though we may not be able to get rid of the term "person" when we want to give a complete description of the world. It is possible that we ascribe experiences to subjects and that we should call these subjects "persons" not "physical" and/or "psychological continuities."

Another way to state what constitutive reductionists hold is this. They claim that what makes different experiences belong to one person is not the fact that they belong to a single separately existing entity. Rather, what makes experiences intrapersonal should be explained in terms of other facts, such as the fact that they are psychologically continuous with one another or the fact that they are associated with a single body.

[2] This is why Parfit calls for a revisionary metaphysics, rather than a descriptive metaphysics: he claims that we have to revise the use of certain of our concepts (see, e.g., Parfit Reasons, ix).

[3] For the term 'constitutive reductionism,' see Parfit "Experiences" and Parfit "Is Personal".

[4] For this reference to Shoemaker, see Parfit "Experiences" (268 n.9).

[5] For a distinction between these two kinds of reductionism, see Parfit "Experiences" (223).

A metaphysical nonreductionist, on the other hand, claims that persons are separately existing entities over and above their physical and psychological continuity. An example of a metaphysical nonreductionist would be someone who identifies persons in accordance with their soul and does not take this soul to be fully constituted by any combination of further entities. This metaphysical nonreductionist could believe in the transmigration of the soul: perhaps she believes that she is identical to some past person from whom her soul has migrated, even though that person's body is not continuous with her current body, that person's character is radically different, and she has no memory of that person's experiences.

Let's return to reductionism. Within constitutive reductionism, there is still one big division to be made. Some reductionists, such as Bernard Williams and Thomas Nagel, argue that a person stays the same person as long as there is a certain degree of physical continuity. Other reductionists, such as Sydney Shoemaker and Parfit, hold that a person stays the same as long as there is a certain degree of nonbranched psychological continuity.

Below, we will look at Parfit's argument for his position. Parfit argues for his view by stating that we should be either nonreductionists or reductionists, by advancing that there is no evidence for the nonreductionist view, and by demonstrating how we can describe psychological continuity in a way that does not presuppose personal identity.

Even when Parfit's argument is considered formally valid, discussion about the truth of his premises and his method is possible.

Two of the premises that could be questioned are premise 6 and premise 7. Can quasi-memories really be called 'memories', or are they only bits of information? If the latter is the case, could quasi-memory then still be said to be an instance of psychological continuity?

As far as Parfit's method is concerned, one could question his appeal to a thought experiment. Parfit imagines a world in which we could have memories of experiencing an event at which we were in fact not present. Philosophers develop thought experiments like these to become clear on our intuitions about a certain concept. They ask something like "If x were the case, what would we then think about A?" There is controversy over whether it is legitimate to appeal to thought experiments in philosophical arguments. Some philosophers, such as Quine ("Identity," 490) and Wittgenstein (*Zettel*, proposition 350), claim that doing so would mean that we attribute a power to words which they in fact do not have. They argue that, being in this world, we cannot really predict what our attitudes in another world would be. They also question what our attitudes in a world unlike ours could possibly say about our attitudes in the world in which we actually live.

> We are not separately existing entities, apart from our brains and bodies, and various interrelated physical and mental events. Our existence just

> involves the existence of our brains and bodies, and the doing of our deeds, and the thinking of our thoughts, and the occurrence of certain other physical and mental events. Our identity over time just involves (a) Relation R – psychological connectedness and/or psychological continuity – with the right kind of cause, provided (b) that this relation does not take a 'branching' form, holding between one person and two different future people. (Parfit *Reasons*, 216)

Defining Premises

P1. When we ask what persons are, and how they continue to exist, the fundamental choice is between two views: the nonreductionist view and the reductionist view (Parfit *Reasons*, 273).

P2. "On the non-reductionist view, persons are separately existing entities, distinct from their brain and bodies and their experiences" (ibid., 275). On this view, persons are entities whose existence must be all-or-nothing (cf., ibid., 273).

P3. On the reductionist view, "persons exist. And a person is distinct from his brain and body, and his experiences. But persons are not separately existing entities. The existence of a person, during any period, just consists in the existence of his brain and body, the thinking of his thoughts, the doing of his deeds, and the occurrence of many other physical and mental events" (cf., ibid.).

Arguments in Defense of the Reductionist View

P4. The reductionist view is true (A) if the occurrence of psychological continuity does not presuppose that a person holds these psychological events together and (B) if we should reject the belief that persons are separately existing entities.

A. The occurrence of psychological continuity does not presuppose that a person holds these psychological events together.

P5. We could think of memories as instantiations of quasi-memories.

P6. I would have an "accurate quasi-memory of past experience if I seem to remember having an experience; someone did have this experience; and my apparent memory is causally dependent on that past experience" (ibid., 220). An example of my quasi-memory of another person's past experience could be this: this person experiences something; a memory of this experience is formed; this memory gets stored on some device and is then downloaded to my brain.

P7. The continuity of quasi-memory is an instantiation of psychological continuity. Or, in other words: if there is continuity of quasi-memory (P(x)), then there is an instantiation of psychological continuity (Q(x)). Formalized, this gives: ($\forall x$(P(x) $\rightarrow$ Q(x)).

P8. If we were aware that our quasi-memories may be of other people's past experiences, as well as of ours, these quasi-memories would and should not be automatically combined with the belief that these memories are about our own experiences. In logical language, this means that the continuity of quasi-memory (P) is consistent with the idea that this continuity can be shared by different persons (R). This relationship of consistency can be formalized as: $\exists x$(P(x) & R(x)).

C1. A certain continuity of quasi-memory can be shared by different persons. Or: P(a) & R(a) (elimination of the existential quantifier, P8).

C2. There is continuity of quasi-memory (P(a)) (simplification, C1).

C3. The occurrence of a certain continuity of quasi-memory implies the occurrence of a certain psychological continuity: P(a) $\rightarrow$ Q(a) (elimination of the universal quantifier, P7).

C4. There is an instantiation of psychological continuity (Q(a)) (*modus ponens*, C2, C3).

C5. Something has the property of being shared by different persons (R(a)) (simplification, C1).

C6. The property of being psychologically continuous is consistent with the property of being shared: Q(a) & R(a) (conjunction,C4, C5).

C7. Psychological continuity is consistent with this continuity not being shared by different persons: $\exists x$(Q(x) & R(x)). Or, in other words: the occurrence of psychological continuity does not presuppose that one person holds these psychological events together) (introduction of the existential quantifier, C6).

B. We should reject the belief that persons are separately existing entities.

P9. If we do not have evidence for the claim that persons exist as separately existing entities, then we should reject this belief (ibid., 224).

P10. We do not have any awareness of the continued existence of a separately existing subject.

P11. We do not have "evidence for the fact that psychological continuity depends chiefly, not on the continuity of the brain, but on the continuity of some other entity" (ibid., 228).

P12. We do not have good evidence for the belief in reincarnation (ibid.). Neither do we have evidence for the existence of Cartesian egos (i.e., thinking nonmaterial substances); it seems like they are neither "publicly observable" nor "privately introspectible facts" (ibid.).

P13. There are no other reasons than the ones in P10, P11, and P12 to believe in the existence of a separately existing subject of experiences.
C5. We have no evidence for the claim that we are separately existing entities (P10, P11, P12, P13).
C6. We should reject the belief that persons exist as separately existing entities (*modus ponens*, P9, C5).
C7. The reductionist view is true (*modus ponens*, P4, C1, C6).

21

Split-Case Arguments about Personal Identity

Ludger Jansen

Parfit, Derek. *Reasons and Persons*. Oxford: Oxford University Press, 1984.
Shoemaker, Sydney, and Richard Swinburne. *Personal Identity (Great Debates in Philosophy)*. Oxford: Blackwell 1984.

In the empiricist tradition, it is a common move to account for the diachronic identity of a person in terms of shared mental properties or continuity of memories (e.g., Locke) or in terms of shared matter, especially of the brain. But all these criteria allow for "split cases," that is, for two or more candidates fulfilling the requirements, which cause trouble with the formal properties of the identity relation (i.e., reflexivity, symmetry, and transitivity). For example, a brain can be divided and both halves implanted in different bodies: which of these, if any, is the same person as the original one? Two individuals could even share most of their memories – but this does not make them the same person. Thus, none of these criteria can be the decisive factor for personal identity. Some philosophers, such as Richard Swinburne (#24), argue for dualism and conclude that there must be some immaterial factor, the soul, that accounts for personal identity. Others, such as Derek Parfit, conclude that we should discard the concept of personal identity altogether and rather replace it with a nonsymmetric successor relation that allows for such split cases.

Just the Arguments: 100 of the Most Important Arguments in Western Philosophy,
First Edition. Edited by Michael Bruce and Steven Barbone.

> There are no logical difficulties in supposing that we could transplant one of [a person] P_1's [brain] hemispheres into the skull from which a brain had been removed, and the other hemisphere into another such skull, and that both transplants should take, and it may well be practically possible to do so. […] If these transplants took, clearly each of the resulting persons would behave to some extent like P_1, and indeed both would probably have some of the apparent memories of P_1. Each of the resulting persons would then be good candidates for being P_1. After all, if one of P_1's hemispheres had been destroyed and the other remained intact and untransplanted, and the resulting person continued to behave and make memory claims somewhat like those of P_1, we would have had little hesitation in declaring that person to be P_1. The same applies, whichever hemisphere was preserved […]. But if it is, that other person will be just as good a candidate for being P_1. […] But […] that cannot be – since the two persons are not identical with each other. (Shoemaker and Swinburne, 15)

P1. A_1 and A_2 are two distinct persons.
P2. At $t_2 > t_1$, A_1 and A_2 are such that each of A_1 and A_2 share exactly the same amount of the X that A had at t_1.
P3. X is the decisive factor for personal identity (e.g., body mass, brain mass, memories, character traces), that is, for any persons A_1 and A_2 and any times t_1 and t_2, if A_2 has at t_2 most of the X that A_1 had at t_1, then A_1 and A_2 are the same person (assumption for *reductio*).
  C1. A_1 is the same person as A (*modus ponens*, P3, P2).
  C2. A_2 is the same person as A (*modus ponens*, P3, P2).
P4. If X is the same person as Y, then Y is the same person as X (symmetry of identity).
  C3. A is the same person as A_2 (*modus ponens*, P4, C2).
P5. If A_1 is the same person as A and A is the same person as A_2, then A is the same person as A_2 (transitivity, C1, C3).
  C4. A_1 is the same person as A_2 (*modus ponens*, conjunction, P5, C1, C3).
  C5. No such X can be the decisive factor for personal identity (*reductio*, P1–C4).

22

The Ship of Theseus

Ludger Jansen

Hobbes, Thomas. "*De corpore*," in *The English Works of Thomas Hobbes*, Vol. 1, edited by Sir William Molesworth. London: John Bohn, 1839.
Plato. *Phaedo*, in *Five Dialogues*, 2nd edn., translated by G. M. A. Grube, revised by J. M. Cooper, 93–154. Indianapolis: Hackett, 2002.
Plutarch. "Life of Theseus," in *Lives*, translated by Bernadotte Perrin, vol. I, 1–87. Cambridge, MA: Harvard University Press, 1967.

The "Ship of Theseus" is an intriguing puzzle about identity through time. It is based on the custom of the Athenians to send Theseus' ship each year on a sacred voyage to Delos, because it was believed that Apollo once saved the lives of Theseus and his fourteen fellow-travellers. The ritual was annually repeated for a long time, and hence the ship needed continual repair, new planks being substituted for the old ones. Plutarch relates to us that already the Athenian philosophers had discussed whether the ship is still the same ship although it consists, after a while, entirely of new planks (Plutarch, "Life of Theseus" §22–3; cf., Plato, *Phaedo* 58a–c). Hobbes put a sophisticated twist to the story: Suppose, he said, that someone collected the old planks and put them together again in the end, thus restoring the old ship. The same ship, then, seems to exist twice, which is absurd. Hobbes used this argument to support his version of relative identity: the original ship T1 and the restored ship T2 share the same matter, whereas the original ship and the repaired ship T3 share the same form.

Just the Arguments: 100 of the Most Important Arguments in Western Philosophy, First Edition. Edited by Michael Bruce and Steven Barbone.

> [I]f, for example, that ship of Theseus, concerning the difference whereof made by continual reparation in taking out the old planks and putting in new, the sophisters of Athens were wont to dispute, were, after all the planks were changed, the same numerical ship it was at the beginning; and if some man had kept the old planks as they were taken out, and by putting them afterwards together in the same order, had again made a ship of them, this, without doubt, had also been the same numerical ship with that which was at the beginning; and so there would have been two ships numerically the same, which is absurd. (Hobbes Chapter 11, 136)

P1. T1 is identical with T2.
P2. It is not the case that T2 is identical with T3.
P3. T3 is identical with T1 (assumption for *reductio*).
C1. T3 is identical with T2 (transitivity of identity, P1, P3).
C2. T2 is identical with T3 (symmetry of identity, C1).
C3. It is not the case that T2 is identical with T3 and T2 is identical with T3 (conjunction, P2, C2).
C4. It is not the case that T3 is identical with T1 (r*eductio*, P3–C3).

25

Two Arguments for the Harmlessness of Death

Epicurus' Death is Nothing to Us Argument

Steven Luper

Epicurus. "Letter to Menoeceus," in *Greek and Roman Philosophy after Aristotle*, edited by Jason Saunders, 49–52. New York: The Free Press, 1966.

Epicurus (341–270 BCE) is most famous for arguing that death is nothing to us. His position is still discussed today, partly because it is not immediately clear where his argument fails and partly because the implications of his conclusion would be important. For example, it seems to follow that we have no reason to avoid death and also that if we save people from death, we are not doing them any good. If death is not bad for us, it seems, living is not good for us.

Epicurus makes his argument in the course of defending a more substantial thesis, namely that anyone can achieve, and then maintain, *ataraxia*, or perfect equanimity. The achievement of complete equanimity requires so situating ourselves that nothing will harm us, so that we have nothing to dread. Since death appears to be harmful indeed, and hence something that a reasonable person will dread, Epicurus needed to explain why it is not.

His argument can be found in the following passage, taken from his "Letter to Menoeceus":

> Death […], the most awful of evils, is nothing to us, seeing that, when we are, death is not come, and, when death is come, we are not. (50)

Unfortunately, it is not clear that this argument accomplishes what Epicurus wanted it to do. The problem is that the term 'death' might mean

Just the Arguments: 100 of the Most Important Arguments in Western Philosophy, First Edition. Edited by Michael Bruce and Steven Barbone.

at least two different things. First, it might signify an event: our ceasing to live. Call this "dying." Second, it might signify a state of affairs: the state of affairs we are in as a result of our ceasing to live. Call this "death." Both dying and death appear to harm us, and hence both threaten our equanimity. But Epicurus' argument shows, at best, that death is nothing to us.

This argument is directed at death rather than dying, but it is possible to substitute 'dying' for 'death'.

P1. We are not affected by an event or state of affairs before it happens.
P2. Death is an event or state of affairs.
C1. Death does not affect us before it happens (instantiation, P1, P2).
P3. If death affects us while we are alive, it affects us before it happens.
C2. Death does not affect us while we are alive (*modus tollens*, P3, C1).
P4. If death affects us while we are dead, it affects us when we do not exist.
P5. We are not affected by anything when we do not exist.
C3. We are not affected by death when we do not exist (instantiation, P5).
C4. Death does not affect us while we are dead (*modus tollens*, P4, C3).
C5. It is not the case that death affects us while we are alive or while we are dead (conjunction, C2, C4).
P6. If death affects us, it affects us while we are alive or while we are dead.
C6. Death does not affect us (*modus tollens*, P6, C5).
P7. What does not affect us is nothing to us.
C7. Death is nothing to us (*modus ponens*, P7, C6).

It is possible to substitute 'dying' for 'death' in this argument, but the resulting argument will clearly be unsound. The problem, of course, is P6, which can easily be challenged on the grounds that dying can affect us while we are dying.

Lucretius' Symmetry Argument

Nicolas Bommarito

Luctretius. *On the Nature of Things*, translated by Martin Ferguson Smith. Indianapolis: Hackett, 2001.
Kaufman, Frederick. "Death and Deprivation; or, Why Lucretius' Symmetry Argument Fails." *Australasian Journal of Philosophy* 74, 2 (1996): 305–12.
Nagel, Thomas. "Death" in *Mortal Questions*. Cambridge: Cambridge University Press, 1997.
Warren, James. *Facing Death*. Oxford: Oxford University Press, 2004.

Symmetry arguments attempt to show the fear of death to be irrational by appeal to similarities between time before our birth and the time after our death. This type of argument has its origin in the philosophy of Epicurus (341–270 BCE), but its most famous statement is in Lucretius' (*c.*99 BCE–*c.*55 BCE) philosophical epic *De Rerum Natura* (*On the Nature of Things*). The scope of the poem is wide, dealing with physics, metaphysics, psychology, and other fields. The clearest statement of the symmetry argument comes near the end of book III:

> Look back now and consider how the bygone ages of eternity that elapsed before our birth were nothing to us. Here, then, is a mirror in which nature shows us the time to come after our death. Do you see anything fearful in it? Do you perceive anything grim? Does it not appear more peaceful than the deepest sleep? (Lucretius III, 972–75)

The argument draws a similarity between pre-natal nonexistence and post-mortem nonexistence; they both are simply states in which we fail to exist. It then notes that we do not fear the time before our birth in which we did not exist, so the time after our death warrants a similar attitude. It is important to remember that the argument is about the fear of death (the state of nonexistence), not the fear of dying (the process of going out of existence).

There are several criticisms of this kind of argument. Thomas Nagel suggests that post-mortem nonexistence is a deprivation in a way that pre-natal nonexistence is not; one who dies is robbed of life in a way that those yet to be conceived are not. Someone whose watch has just been stolen is not in the same state as someone who never owned a watch; they are both watch-less, but one of them has lost something. One might also think that fear itself has a temporal aspect and is essentially future-directed in the way it is natural to fear being fired next week but not to fear having been fired last week.

Another response to the argument is to grant the symmetry, but use our fear of death as a premise rather than our lack of fear of the time before we existed. Another way to have similar attitudes toward both states is to fear both the time before we existed and the time after our death.

P1. The pre-natal state is a kind of nonexistence.
P2. The post-mortem state is a kind of nonexistence.
  C1. Pre-natal and post-mortem states are relevantly similar; both are states of nonexistence (conjunction, P1, P2).
P3. If states are relevantly similar, then they warrant similar attitudes.
  C2. The pre-natal and post-mortem states warrant similar attitudes (*modus ponens*, C1, P3).
P4. The pre-natal state does not warrant fear.
  C3. Post-mortem nonexistence does not warrant fear (instantiation, C2, P4).

30

An Argument for Free Will

Gerald Harrison

Clarke, Randolph. "Toward a Credible Agent-Causal Account of Free Will." *Noûs* 27 (1993): 191–203.

van Inwagen, Peter. *An Essay on Free Will*. Oxford: Oxford University Press, 1983.

____. "How to Think about the Problem of Free Will." *Journal of Ethics* 12 (2008): 327–41.

Reid, Thomas. *Essays on the Active Powers of the Human Mind*. Cambridge, MA: The MIT Press, 1969.

Strawson, Peter F. "Freedom and Resentment." *Proceedings of the British Academy* 48 (1962): 1–25.

Some philosophers think that our decisions are free only if uncaused, others that causation is needed to prevent our decisions being uncontrolled; some think that the causation needs to be indeterministic, others that it needs to be deterministic, and others that it does not matter either way.

Nevertheless, there is near unanimous agreement that free will is needed to ground moral responsibility. That is to say, free will is required if we are to deserve praise, blame, reward, or punishment for our deeds, and if a host of so-called "reactive attitudes" such as resentment, guilt, and forgiveness are appropriate.

This common ground among disputants provides the basis for a positive argument for free will. Versions of this argument (which has no specific

Just the Arguments: 100 of the Most Important Arguments in Western Philosophy, First Edition. Edited by Michael Bruce and Steven Barbone.

name) have been presented by Thomas Reid, Randolph Clarke, Peter van Inwagen (*Essay*), and Peter Strawson, among others.

Just as it is widely agreed that moral responsibility requires free will, it is also widely agreed that we are morally responsible for at least some of what we do some of the time. For Reid, it was a first principle "that some aspects of human conduct deserve praise, others blame" (361). According to Peter Strawson, our commitment to moral responsibility is so deeply rooted that it is simply inconceivable that we could give it up, and thus the reality of moral responsibility sets a boundary condition for where rational argument can lead.

If our moral responsibility is beyond reasonable doubt, then it must be beyond reasonable doubt that we possess free will, as the former presupposes the latter. Thus, we get our positive argument for free will.

Not everyone accepts this argument. A significant minority of philosophers deny that we are morally responsible. There are, after all, powerful arguments both for thinking that free will is incompatible with determinism and for thinking that it is incompatible with indeterminism. Such arguments can be used to raise doubts about whether we have free will, and so to raise doubts about moral responsibility.

For most, however, the belief that we are morally responsible has greater initial plausibility than any of the premises of an argument leading to the denial of free will. Moral responsibility therefore provides the best positive argument for thinking that we do have free will.

> There are, moreover, seemingly unanswerable arguments that, if they are correct, demonstrate that the existence of moral responsibility entails the existence of free will, and, therefore, if free will does not exist, moral responsibility does not exist either. It is, however, evident that moral responsibility does exist: if there were no such thing as moral responsibility nothing would be anyone's fault, and it is evident that there are states of affairs to which one can point and say, correctly, to certain people: That's *your* fault. (van Inwagen "How to Think")

P1. If we are morally responsible then we have free will.
P2. We are morally responsible.
C1. We have free will (*modus ponens*, P1, P2).

31

Frankfurt's Refutation of the Principle of Alternative Possibilities

Gerald Harrison

Frankfurt, Harry. "Alternate Possibilities and Moral Responsibility." *Journal of Philosophy* 45 (1969): 829–39.

Fischer, John M. "Frankfurt-Style Compatibilism," in *Free Will*, edited by Gary Watson, 190–211. Oxford: Oxford University Press, 2003.

Widerker, David, and Michael McKenna, (eds.). *Moral Responsibility and Alternative Possibilities*. Farnham, UK: Ashgate, 2006.

Endorsed by Aristotle, Hume, Kant, and many others the "Principle of Alternative Possibilities" (PAP for short) states:

> PAP: A person is morally responsible for what she has done only if she could have done otherwise.

Historically, PAP has been one of the most popular routes to "incompatibilism" about moral responsibility (incompatibilism is the view that moral responsibility and causal determinism – the thesis that there is only one future compatible with the past and the laws of nature – are incompatible). After all, if determinism is true, there's a sense in which no one could ever have acted differently. "Compatibilists" (those who believe determinism and moral responsibility to be compatible) resisted this argument by arguing that PAP should be given a controversial "conditional" interpretation according to which an agent could have done otherwise if he would have done so had he desired.

Just the Arguments: 100 of the Most Important Arguments in Western Philosophy,
First Edition. Edited by Michael Bruce and Steven Barbone.

But in 1969, the philosopher Harry Frankfurt devised an argument to refute PAP. Frankfurt argued that it is possible for circumstances to arise in which it is clear that a person could not have done otherwise yet also clear that he is morally responsible for his deed. The defining feature of what has now become known as a "Frankfurt-style case" is that an intervention device does not intervene in a process leading to an action but would have intervened if the agent had been about to decide differently. The presence of the intervention mechanism rules out the possibility of the agent's deciding differently, yet because the intervention mechanism plays no role in the agent's deliberations and subsequent action, it seems clear that the agent is fully morally responsible for his action; hence PAP is refuted.

By refuting PAP, Frankfurt's argument closes off one of the major routes to incompatibilism and allows compatibilists to bypass the debate over the correct interpretation of PAP.

Frankfurt's argument remains the focus of considerable debate, with detractors arguing that it is impossible to construct a Frankfurt-style case in which all relevant alternative possibilities have been expunged.

> Suppose someone, Black, let us say wants Jones to perform a certain action. Black is prepared to go to considerable lengths to get his way, but he prefers to avoid showing his hand unnecessarily. So he waits until Jones is about to make up his mind what to do, and he does nothing unless it is clear to him (Black is an excellent judge of such things) that Jones is going to decide to do something other than what he wants him to do. If it does become clear that Jones is going to decide to do something else, Black takes effective steps to ensure that Jones decides to do, and that he does do, what he wants him to do. Whatever Jones's initial preferences and inclinations, then, Black will have his way […].
>
> Now suppose that Black never has to show his hand because Jones, for reasons of his own, decides to perform and does perform the very action Black wants him to perform. In that case, it seems clear, Jones will bear precisely the same moral responsibility for what he does as he would have borne if Black had not been ready to take steps to ensure that he do it. (Frankfurt, 835–6)

P1. An agent is morally responsible for what he has done only if he could have done otherwise (PAP).
P2. If PAP is true, then a Frankfurt-style case will absolve its subject from moral responsibility.
P3. Frankfurt-style cases do not absolve their subjects from moral responsibility.
C1. PAP is false (*modus tollens*, P2, P3).

32

Van Inwagen's Consequence Argument against Compatibilism

Grant Sterling

van Inwagen, Peter. *An Essay on Free Will*. Oxford: Clarendon Press, 1983.

One of the most famous recent arguments in the free will and determinism debate is Peter van Inwagen's consequence argument, which aims to show that compatibilism is false. Compatibilism is the view that all our actions could be fully determined by the laws of physics and yet at the same time we could have free will in the sense necessary for moral responsibility. Van Inwagen introduces the essence of this argument near the beginning of his book on free will and then goes on to gives three detailed technical versions of the argument. Included here is the simple version and the first technical formalization (which aims to show that under determinism we could never act in any way other than the way in which we do act).

> If determinism is true, then our acts are consequences of the laws of nature and events in the remote past. But it is not up to us what went on before we were born, and neither is it up to us what the laws of nature are. Therefore, the consequences of these things (including our present acts) are not up to us. (16)

> Consider any act that (logically) someone might have performed. If it should turn out that this act was incompatible with the state of the world before that person's birth taken together with the laws of nature, then it follows that that person could not have performed that act. Moreover, if determinism is true, then just any deviation from the actual course of events would be incompatible with any past state of the world taken together with the laws of nature. Therefore, if determinism is true, it never has been within

Just the Arguments: 100 of the Most Important Arguments in Western Philosophy,
First Edition. Edited by Michael Bruce and Steven Barbone.

> my power to deviate from the actual course of events that has constituted my history. (75)

P1. If determinism is true, then our acts are consequences of the laws of nature and events in the remote past.
P2. The laws of nature and events in the remote past are not up to us.
P3. If something is not up to us, then its consequences are not up to us.
  C1. If the laws of nature and events in the remote past are not up to us, then their consequences are not up to us (substitution, P2, P3).
  C2. Consequences of the laws of nature and events in the remote past are not up to us (*modus ponens*, P2, C1).
  C3. If determinism is true, then our acts are not up to us (in our control, within our ability) (substitution, C2, P1).
P4. If our acts are not up to us, then we're not responsible for them.
  C4. If determinism is true, we're not responsible for any of our acts (hypothetical syllogism, C3, P4).

Van Inwagen's First Formalization

Definitions:

> Let 'U' be a complete description of the state of the universe right now.
> Let 'U – 1' be a complete description of the state of the universe the day before some person 'X' was born.
> Let 'A' be some action that X did *not* perform.
> Let 'L' be the laws of nature.

P1. X cannot change U – 1 (no one can change the past state of the universe at a time before she was even born).
P2. X cannot change L (no one can change the laws of nature).
P3. If determinism is true, then {(U – 1 plus L), entails U} (follows from the concept of determinism).
P4. If X had done A, then not-U (A is an action that didn't occur, so if it had occurred the universe wouldn't be exactly the same as it is now).
  C1. If X could have done A, X could have made U false (follows semantically from P4).
  C2. If X could have made U false, then X could have made (U – 1 plus L) false (transposition, P3).
  C3. If X could have made (U – 1 plus L) false, then X could have made L false (De Morgan's, C2, P1, and disjunctive syllogism).
  C4. X could not have made L false (P2).
  C5. X could not do A (*modus tollens*, C3, C4, and a series of implicit hypothetical syllogisms).

33

Fatalism

Fernando Migura and Agustin Arrieta

Aquinas, Thomas. *Summa Theologiae*, translated by Fathers of the English Dominican Province, *The Summa Theologiae*, 2nd rev. edn., 22 vols. London: Burns, Oates & Washbourne, 1912–36. Reprinted in 5 vols., Westminster: Christian Classics, 1981. E-text in HTML available at www.newadvent.org/summa

Aristotle. *Aristotle Categories and De Interpretatione*, translated with notes and glossary by J. L. Ackrill. Oxford: Clarendon Press, 1961.

Augustine, Saint. *On Free Choice of the Will*, translated, with introduction by Thomas Williams. Indianapolis: Hackett, 1993.

Rice, Hugh. "Fatalism." *The Stanford Encyclopedia of Philosophy* (Fall 2009 edn.), edited by Edward N. Zalta, available at http://plato.stanford.edu/archives/fall2009/entries/fatalism

According to the philosophical doctrine called "fatalism," everything that happens does so inevitably. Suppose that something is going to happen tomorrow; let's say that it is going to rain. If it is true now that tomorrow it is going to rain, then it can't be true that it won't rain tomorrow, so it is necessary to rain tomorrow. On the other hand, if it is false now that tomorrow it is going to rain, then it can't be true that it will rain tomorrow, so it is impossible to rain tomorrow; that is, it is necessary that it won't rain tomorrow. Since the same reasoning can be applied to every event, everything that happens does so necessarily and inevitably.

Just the Arguments: 100 of the Most Important Arguments in Western Philosophy,
First Edition. Edited by Michael Bruce and Steven Barbone.

Let us see the structure of the argument from which fatalism is concluded. Let *p* be: "It is going to rain tomorrow" (or whatever declarative sentence that describes an event that you think that can happen tomorrow). Then the argument has the following structure:

P1. If it is true now that *p*, then necessarily *p*.
P2. If it is true now that not *p*, then necessarily not *p*.
P3. It is true now that *p* or it is true now that not *p*.
C1. Necessarily *p* or necessarily not *p* (constructive dilemma, P1, P2, P3).

This argument is unsound because it is clear that the conclusion is false, but it is not so clear where it goes wrong. The classical solution has to do with a known ambiguity (amphiboly) associated with conditional sentences of the form: "If X, then, necessarily Y." This can be interpreted as (a) "It is a necessary truth that if X, then Y" or as (b) "If X, then it is a necessary truth that Y." On the one hand, if premises 1 and 2 are read as (a), they are clearly true but, then, the conclusion doesn't follow from premises. On the other hand, if premises 1 and 2 are interpreted as (b), the conclusion does follow from them, but they presuppose fatalism. So, either the argument is not logically valid or it begs the question.

The first and best known argumentative version of fatalism can be found in the sea-battle argument formulated by Aristotle in Chapter IX of *On Interpretation* (*Peri Hermeneias*, also *De Interpretatione*):

> For if every affirmation or negation is true or false it is necessary for everything either to be the case or not to be the case. For if one person says that something will be and another denies this same thing, it is clearly necessary for one of them to be saying what is true – if every affirmation is true or false; for both will not be the case together under such circumstances. […] It follows that nothing either is or is not happening, or will be or will not be, by chance or as chance has it, but everything of necessity and not as chance has it (since either he who says or he who denies is saying what it is true).
>
> I mean, for example: it is necessary for there to be or not to be a sea-battle tomorrow, but it is not necessary for a sea-battle to take place tomorrow, not for one not to take place – though it is necessary for one to take place or not to take place. (Aristotle *On Interpretation*, IX 18a34, 19a23)

But there are also other known formulations due to St. Augustine and Thomas Aquinas relating to the associated problem of free will. St. Augustine in *On Free Choice of the Will* (Book Three), considers an argument that could be paraphrased as follows:

> If God foreknows that Pope Benedict XVI will sin tomorrow, then necessarily Pope Benedict XVI will sin tomorrow. God foreknows that Pope

Benedict XVI will sin tomorrow. So necessarily Pope Benedict XVI will sin tomorrow.

Another example of this is Thomas Aquinas' discussion of the argument that God's Providence (*Summa Theologiae*, First Part, Question 22) implies fatalism. The argument is built from a supposition like this: During the Creation, God foresaw everything, including, for example, Pope Benedict XVI sinning tomorrow. So, necessarily Pope Benedict XVI will sin tomorrow.

Assuming that what God foreknows or sees is always true, these versions of fatalist arguments are essentially analyzed in the same way. Both arguments count as *modus ponens*: "If X, then, necessarily Y, and X, so, necessarily Y." In both cases, the key issue has to do with the correct interpretation of conditional sentence properly understood as "It is necessarily true that X, then Y."

Let us consider a more familiar example:

(e) "If I know George Clooney is a bachelor, then necessarily George Clooney is unmarried."

Given that I know George Clooney remains Hollywood's most famous bachelor today (September 1, 2010), if I don't interpret correctly the conditional, I can conclude by *modus ponens*, "Necessarily, George Clooney is unmarried." But this conclusion would be equivalent to saying, "There are no possible circumstances in which George Clooney is married," and so a strong conclusion is not justified by the premises. Obviously the correct interpretation of (e) is, "Necessarily, if I know George Clooney is a bachelor, then George Clooney is unmarried."

One of the most known practical consequences of fatalism has to do with the uselessness of decision-making. If someone assumes fatalism, why should she bother making decisions if the outcome is already fixed? This direct consequence of fatalism is clearly illustrated in the famous "lazy argument." For instance, if you feel sick now, it is true now that you will either recover or it is now true that you will die. In any case, by direct application of the fatalist argument, necessarily you recover from your illness or necessarily you die because of it. So, why should you call the doctor or do anything at all? (As is easy to see, this argument has the form of a dilemma too.)

Aristotle was entirely aware of this consequence of fatalism when he said that if everything is and happens of necessity, there would be no need to deliberate or to take trouble thinking that if we do this, this will happen, but if we do not, it will not (see *On Interpretation*, IX 18b26).

34

Sartre's Argument for Freedom

Jeffrey Gordon

Sartre, Jean-Paul. *Being and Nothingness*, translated by Hazel Barnes. New York: Philosophical Library, 1956.

Sartre's argument for freedom is unique in the history of philosophy because it treats freedom as the essential characteristic of human consciousness as opposed to a property or capacity of consciousness or mind. In one of Sartre's famous formulations, "Man is freedom," the idea is that consciousness has no properties at all, that it is nothing more than a relation to real existent things, and it relates to those things by defining their significance. The conscious person must interpret the significance of the existent thing; he must construct a coherent world from what is given. The given has no meaning in itself; whatever meaning it will have derives from the agent's interpretation. For a given state of affairs to function as a cause of my conduct, I must first confer upon that state of affairs a certain meaning, which in turn informs that situation with its power to cause. I, then, am the source of its causal efficacy. But determinism requires that the nature and compelling power of the cause exist in themselves, quite independently of any characteristic of the entity undergoing the cause–effect process. Since this necessary condition of determinism is never met by consciousness, determinism is inapplicable to human experience. Experience cannot be caused. To experience is to appropriate, to interiorize the given, to make it

Just the Arguments: 100 of the Most Important Arguments in Western Philosophy, First Edition. Edited by Michael Bruce and Steven Barbone.

one's own. In virtue of the relationship between consciousness and the given, my freedom to choose is inescapable. Sartre therefore concludes, "Man is condemned to be free" (439).

Suppose that a boy is born into poverty; that is, the socioeconomic condition of his family is much lower than the average. (The idea of poverty, fraught with connotations of disvalue, already presupposes an interpretation.) Trying to explain his later extraordinary drive, we might well cite this early circumstance as formative – indeed, as determinative. But Sartre would insist that such an explanation is quite misleading. The poverty could not have had this effect had the young boy not understood the condition as shameful. Had he thought of it instead as the source of the strong mutual dependency in his family and their consequent bonds of solidarity, the drive for wealth might very well have seemed to him an empty pursuit. Sartre's point would be that a given socioeconomic circumstance must await the interpretation of consciousness before it could function as a cause. Life circumstances cannot impel an effect without the assent of consciousness. Always to have to interpret the given, to have to forge of the given a motive and cause, is the inescapable condition of consciousness. The uncaused source of its own actions, the human being is irremediably free.

> No factual state whatever it may be (the political and economic structure of society, the psychological "state," etc.) is capable by itself of motivating any act whatsoever. For an act is the projection of [consciousness] toward what it is not, and what is can in no way determine by itself what is not. [. . .] This implies for consciousness the permanent possibility of effecting a rupture with its own past, of wrenching itself away from its past so as to be able to consider it in the light of a non-being and so as to be able to confer on it the meaning which it has in terms of the project of a meaning it *does not have*. Under no circumstances can the past in any way by itself produce an act [. . .]. In fact as soon as one attributes to consciousness this negative power with respect to the world and itself [. . .] we must recognize that the indispensable and fundamental condition of all action is the freedom of the acting being. (436)

P1. In order for a given state of affairs deterministically to cause a human action, the causal efficacy of that state of affairs would have to derive exclusively from characteristics of that state of affairs.
P2. A given state of affairs has no meaning in itself.
P3. If a given state of affairs has no meaning in itself, then its meaning must be conferred upon it by the person experiencing it.
  C1. The meaning of a given state of affairs must be conferred upon it by the person experiencing it (*modus ponens*, P2, P3).
P4. The meaning of the state of affairs is the source of its power to motivate (or cause) the action.

P5. If the meaning of the state of affairs is the source of its power to motivate (or cause) the action, then in the case of human action, the causal efficacy of the state of affairs does not derive exclusively from characteristics of that state of affairs.

C2. In the case of human action, the causal efficacy of the state of affairs does not derive exclusively from characteristics of that state of affairs (*modus ponens*, P4, P5).

C3. No given state of affairs can deterministically cause a human action (*modus tollens*, P1, C3).

P6. If no given state of affairs can deterministically cause a human action, then one's actions are free.

C4. Human beings are inescapably free (*modus ponens*, C3, P6).