***Introduction to Philosophy: Classical and Contemporary Readings* (9th Edition)**

**Part IV Summary: Minds, Bodies, and Persons**

**David M. Armstrong, “The Nature of Mind”**

Armstrong’s task in this article is to sketch a theory of the mind that is consistent with a view he calls “materialism.” Materialism, as Armstrong understands it, is just the view that we can give a complete account of human beings in purely physico-chemical terms. We don’t need to make reference to any sort of soul or spirit or immaterial mental properties to make sense of humans. Armstrong contends that our best science is leading us in this direction; thus, it is worth attempting to construct a theory of the mind that is consistent with materialism.

The theory that Armstrong presents is actually a synthesis of two different views: dualism and behaviorism. Dualism is just the Cartesian view that the inner is utterly different than the outer and that there are mental occurrences in just the same sense as there are physical occurrences. Behaviorism, on the other hand, is the view that all talk of mental occurrences can be reduced to talk about outward behavior. Armstrong tries to bring these two views together.

Behaviorism is certainly attractive to materialists, but Armstrong thinks it is open to fatal objections. Most important, it seems to completely overlook the fact that there seems to be a lot going on inside us even when we aren’t displaying any outward behavior. Behaviorists attempt to meet this objection by talking about *dispositions* to behave in certain ways, but Armstrong thinks that even talk of dispositions isn’t robust enough to do justice to what goes on inside our heads. However, he thinks that the behaviorists are right to say that the mind is somehow tied to behavior. Indeed, Armstrong suggests that perhaps the mind is just whatever it is that *causes* outward behavior.

This suggestion is consistent with materialism (unlike dualism), and it does justice to our own experience of our thoughts (unlike behaviorism). Armstrong then goes on to develop this suggestion a bit more, including defining a “mental state” as “a state of the person apt for producing certain ranges of behavior.” Finally, he considers an objection to the account, namely, that the account doesn’t seem to make room for *consciousness*. Surely when we act it’s not only the case that some of our inner states are causing our outward behavior—it’s also true that we are aware that this is going on. So where does consciousness fit in?

Armstrong concludes by suggesting a response to this objection. He draws an analogy between consciousness and perception and then contends that consciousness is nothing but perception or awareness of the state of one’s own mind. And just as the account doesn’t rule out perception in general, so it doesn’t rule out perception that is directed at one’s own mind.

In the end, Armstrong points out that there is much more to be done to develop this sort of account, but he is pleased that it at least offers a promising alternative for those who accept materialism.

**Paul M. Churchland, “Eliminative Materialism”**

Eliminative materialism is the view that our commonsense psychological framework is false and misleading and thus should be jettisoned. Other materialist theories of the mind, such as the identity theory or behaviorism, attempt a reduction of folk psychology to neuroscience, but eliminative materialism asserts that no such reduction is possible because our folk concepts are hopelessly confused. In this selection, Paul M. Churchland outlines and defends this view.

Churchland begins by pointing out that such a radical view is not altogether unprecedented in the history of ideas. To take just one example, in the past many thought that there was a mysterious substance called “phlogiston” that was involved when something burned. But advances in science showed us that this idea was confused, and eventually everyone agreed that phlogiston does not exist. Churchland thinks a similar fate will come to the concepts of folk psychology.

He provides three arguments in favor of eliminative materialism and three arguments against. The arguments for the view are as follows: First, Churchland points to the widespread explanatory failures of folk psychology in support of the view that it is hopelessly confused. Second, he argues via a sort of induction that most of our other folk theories about scientific phenomena have been eliminated (e.g., the view that the universe rotates around the Earth). And third, he points out that eliminative materialism seems more likely to be true than either the identity theory or behaviorism simply because it does not have the extra burden of attempting a reduction of folk concepts to neuroscientific concepts.

Churchland concludes by considering and responding to three arguments against his view. The first is that one’s introspection reveals the existence of pains and beliefs and other folk psychological items. Churchland replies by pointing out that all observation occurs within a conceptual framework. Part of his position is that we need to get rid of our whole conceptual framework. Second, someone might argue that the eliminative materialist cannot coherently assert his theory, because to assert it meaningfully would be to have a belief that causes him to assert it. But Churchland replies that this objection relies on a contentious view about what makes assertions meaningful. Finally, someone might object that at the very least, the eliminative materialist is exaggerating things. Churchland concedes that this may be true and that we should certainly leave room for the possibility that some folk concepts will be reduced and some will be eliminated. However, he at least wants to make the point that we should take the possibility of elimination very seriously.

**Frank Jackson, “What Mary Didn’t Know”**

In this short piece, Frank Jackson clarifies the knowledge argument against physicalism and responds to three objections that have been raised against the argument by Paul Churchland.

The knowledge argument runs as follows: Imagine that Mary is raised in a completely black-and-white environment and that she learns through lectures on black-and-white TV all the physical facts about the real world. If physicalism is true, then Mary knows all there is to know. But it seems that she doesn’t know all there is to know. In particular, when she gets let out of the room, she will learn what it is like to see red and, hence, learn something new about others’ experiences that she did not know before. Thus, physicalism is false.

Jackson presents three clarifications. First, the argument is not about what Mary can imagine; it’s about what Mary can know. Second, it’s not about the intentionality of knowledge and whether knowledge can be transferred over logical implication. And third, it’s not about her learning more about her own experience—it’s about her coming to know something about the experience of others that she previously didn’t know.

Finally, Jackson responds to three of Paul Churchland’s objections. Churchland’s first objection is that the argument equivocates between knowledge by acquaintance and knowledge by description. Jackson responds that the argument is not about different types of knowledge—it is about the scope of Mary’s knowledge. When she is let out into the real world, she learns something new. Churchland’s second objection is that if the argument works against physicalism, it works against dualism, too, as Mary could have learned all the facts about qualia through a black-and-white TV set, too. But Jackson thinks it is implausible to suppose that facts about qualia could be communicated through TV, whereas it is not implausible to suppose the same in the physicalist case. Churchland’s final objection brings up considerations of imagination, but as Jackson already pointed out, the knowledge argument is not about imagination.

Thus, Jackson concludes that the knowledge argument has not yet been refuted.

**Patricia Churchland, “Neurophilosophy”**

In this article, Patricia Churchland articulates “neurophilosophy,” an approach to philosophy that utilizes scientific discoveries to consider a large range of traditional philosophical questions. Particularly, she focuses on the relevance of neurophilosophy for philosophy of mind, especially the debate between physicalists and dualists. Neurophilosophy, Churchland argues, clearly suggests that dualism is false. Through this discussion, Churchland is led to reflect on philosophy more generally, including why philosophers are often hesitant to draw on science to arrive at philosophical answers.

In her introduction, Churchland defines “neurophilosophy,” and provides a helpful introduction to it. She then, in section 2, turns to the question of how the brain and mind are related. She articulates the dualist and physicalists positions, and then argues that science suggests that dualism is false. She then considers two responses a dualism might make to her neurophilosophical argument. They might argue for a kind of epiphenomenalism, or they might argue that the data she adverts to are actually neutral with respect to this debate. Ultimately, Churchland thinks that much of the resistance to neurophilosophy (and Physicalism) is motivated by philosophical biases toward conceptual analysis and necessary truths, which make conceivability more important than, she argues, it is. In the next section, Churchland relates the history of neurophilosophy, particularly her early work in the field. And then, in section 4, she turns to one of Quine’s dogmas, which takes it that conceptual analysis tells us something about the way the world is. She discusses how worrisome this dogma is and thereby argues for a neurophilosophical approach.

**A. M. Turing, “Computing Machinery and Intelligence”**

In this selection, Turing considers the question “Can machines think?” Or, rather, he replaces that question with another: “Can a machine play the imitation game?” The imitation game involves three players. Player A is a man, player B is a woman, and player C’s job is to figure out which player is the man and which is the woman. Player C can ask any questions that he wants to, and players A and B will respond via typewritten messages. Player C cannot see or touch or hear either of the other two players and must try to determine the sex of the players completely from the way they answer the questions. Turing’s question, then, is whether a machine could take the place of either player A or B and answer questions in such a way that a human interrogator would have about the same success distinguishing the human from the machine as he would distinguishing the man from the woman.

Because computers were just getting off the ground when Turing was writing this piece, he spends quite a bit of time explaining what a digital computer is and how it works. Eventually, he turns his attention back to the question with which he started. Could a machine play the imitation game well enough to fool a human interrogator? Turing then considers potential objections to his using this question to replace the original “Can machines think?” What follows is a representative sample of these objections, along with Turing’s responses.

The first objection Turing considers deals with the results of Gödel’s incompleteness theorem and other similar results. Gödel’s theorem shows that there is a defect in any machine, and this objection points out that because there is no parallel defect in human beings, human beings cannot be machines. Turing responds by pointing out that we have no way to be sure that the same defect is not in human beings just as it is in machines.

Another objection raises the point that although machines might be able to perform tasks, in some sense, they cannot be aware that they are performing tasks. Machines are not conscious. But Turing responds that this attitude would ultimately lead to solipsism, because, strictly speaking, we cannot show that other human beings are conscious either.

A third objection is that there are many things you can’t get a computer to do, such as fall in love or be friendly or make mistakes. But Turing thinks that this objection is the product of some sort of induction from actually observing computers. Because computers that existed at the time were not terribly complex, this sort of objector concluded that no computer could be complex enough to do these tasks. But this sort of induction is certainly unwarranted, according to Turing.

Turing ends the piece with some speculative remarks about the possibility of creating a machine with the mind of a child, and then subjecting it to some sort of education process so that it is actually a learning machine.

**John R. Searle, “Minds, Brains, and Programs”**

Searle’s goal in this selection is to argue against the claims of what he calls strong artificial intelligence (AI). Whereas weak AI is the claim that computers give us a powerful tool to study the mind, strong AI is the view that an appropriately programmed computer *is* a mind. Searle argues against strong AI by utilizing his now-famous Chinese Room thought experiment, which runs as follows.

Searle asks us to imagine that he is put in a room and handed slips of paper with Chinese sentences written on them. The room he is in contains many volumes of books that tell him exactly which symbols to hand back outside the room in response to the various strings of symbols that he is handed. In other words, he is given some input in the form of Chinese sentences, he uses that input to produce an output based on what he finds in the books, and the output is also in the form of Chinese sentences. He asks us to imagine further that he gets very good at figuring out which books he needs to consult so that he can complete the process very quickly. Now Searle asks: Does he understand Chinese?

Searle thinks it is clear that merely being a part of this setup is not sufficient for his understanding Chinese. However, he thinks that according to strong AI, this sort of setup is all that is needed for thinking. Therefore, Searle concludes that strong AI must be mistaken. And, in general, his thought experiment appears to show that there can be two machines that both pass the Turing test, even though only one of those machines (the human, in this case) truly understands what’s going on.

There are various objections to consider at this point. Here are a few. The systems reply objects that even though the human inside of the Chinese room doesn’t understand Chinese, nevertheless the entire system, including books and pieces of paper, does understand Chinese. But Searle thinks that this objection, in addition to being incredible, overlooks the fact that the same thought experiment can be run even if we assume that the human has internalized all of the books and other parts of the system.

The brain simulator reply points out that we could make a computer that exactly parallels whatever goes on in a fluent Chinese speaker’s brain when he or she understands Chinese, and that such a computer would surely understand Chinese. But Searle responds that even in such a case, there would be no understanding, because you could just change the original thought experiment so that the man in the room turns on and shuts off water valves that correspond to the neural firings of a Chinese speaker. In such a case, the man doesn’t understand.

Finally, someone might object that we will probably eventually be able to build machines that process in whatever way is required for understanding. Searle admits that this might be true, but he thinks it trivializes the claims of strong AI, because strong AI is supposed to give us insight into what understanding is all about.

Overall, Searle concludes that although we probably could create a computer with understanding, it won’t be one whose operation is defined as an instantiation of a computer program.

**John Perry, “A Dialogue on Personal Identity and Immortality”**

This selection is an entertaining dialogue in which three characters discuss philosophical issues surrounding personal identity and immortality. It’s a fairly long selection, so this summary includes only the highlights.

To begin with, the characters are Weirob, a philosopher who is on her deathbed; Miller, friend to Weirob and chaplain; and Cohen, a former student of Weirob’s. The main points of the dialogue are brought out by Weirob and Miller, who stand on opposite sides of the theism debate, although Cohen occasionally throws in an insightful comment.

The dialogue begins with Weirob issuing a challenge to Miller. She challenges him to persuade her that survival after bodily death is possible. Miller then proceeds to give various responses to this challenge. He first tries out the idea that Weirob will survive bodily death because there will be someone in heaven who is qualitatively identical to her. But Weirob points out that mere qualitative identity is not the identity relevant to her concern. Miller then tries arguing that because Weirob is really a soul, the soul is what will survive after bodily death. But Weirob responds that we have no way of judging sameness of soul over time, as the soul is something we cannot see. So, if our judgments about personal identity had to do with sameness of soul, we would never know whether we were right in judging that one person at one time is the same at another time. But surely, we do know that our judgments are correct. Miller then tries appealing to sameness of psychological characteristics over time to explain sameness of soul over time, but this falls prey to similar objections. Finally, Miller suggests that he can generalize from his own case, as he clearly knows that there is a correlation between body and soul in his own case. But Weirob questions whether he can be certain about even that.

During the second night of the dialogue, Miller argues directly against explaining personal identity in terms of bodily identity. He points out that we can know that we are the same person over time without ever opening our eyes and, thus, without knowing that we have the same body that we used to. He then goes on to give an account of personal identity in terms of person-stages that are connected via memory over time. He gets himself into a bit of a circle when he tries to explain how to distinguish between someone who actually does remember an event and someone who only seems to remember an event, but Cohen pipes up to help out. Cohen suggests that if we introduce causation into the person-stage memory picture, we can get out of the circle. But then he runs into problems explaining how God could make it the case that a person-stage here on earth stands in a causal relationship to a person-stage up in heaven.

Finally, on the third night, the three characters leave the topic of an afterlife and discuss personal identity itself. They spend most of the rest of the dialogue discussing in detail a brain transplant thought experiment.

In the end, Weirob dies, but at least she had an interesting philosophical discussion before she went.

**Derek Parfit, “Personal Identity”**

In this article, Parfit attempts to undermine two common beliefs about personal identity through time and to propose an alternative account of what matters to us when we speak of personal identity. The two beliefs he aims to undermine can be stated as follows:

1. There is always a definite answer to questions about personal identity.

2. Unless questions about personal identity have a definite answer, we cannot answer certain other important questions about things such as survival, memory, and responsibility.

Against the first belief, Parfit describes a case meant to support the conclusion that the belief is implausible. Against the second, Parfit argues that those other important questions (about survival, memory, and responsibility) only *appear* to presuppose questions about personal identity, and once they are freed from this presupposition, we will see that personal identity is not that important after all.

First, the problem case. Consider a case of fission, where my brain is split and put into two new bodies. What happens to me? It seems there are only three possibilities: (1) I do not survive, (2) I survive as one of the two people, or (3) I survive as both. But each of these has its share of implausibility. The trouble with option 1 is that it is plausible to suppose both that someone can survive having their whole brain transplanted into another body and that someone can survive with half of his or her brain destroyed. If we put those plausible suppositions together, it also seems plausible to suppose that someone can survive having his or her brain split and then transplanted. The trouble with option 2 is that any choice of which person I survive as would be arbitrary. And the trouble with option 3 is that it seems to violate the transitivity of identity (that is, if “survival” is a matter of identity). What do we do?

This problem case suggests that the first belief mentioned earlier is mistaken. Maybe there is no right answer to the question of personal identity here. Moreover, Parfit suggests that this case shows us that we need a sense in which one person can survive as two. If we give up talking about *identity*, we can consistently go with option 3—that I survive as two people without violating the transitivity of identity. And because it seems that what matters to survival is present in this problem case, we should find a sense of “survival” according to which I do survive a case of fission as both resulting people.

To find such a sense, Parfit turns to talking about psychological continuity, as it seems that the relations that matter when we speak of personal identity are all relations of psychological continuity. He develops an account of psychological continuity, and he explains how it can account for our intuitions about a range of cases.

J. **David Velleman, “So It Goes”**

In this piece, Velleman sets out to explore the Buddhist idea that the enduring self is an illusion. He agrees with Derek Parfit that giving up the idea of an enduring self can have beneficial consequences but thinks that its consequences are even more radical than Parfit imagined.

Velleman begins by exploring the very idea of an enduring self, arguing that we get the idea of an enduring self in the first place by projecting our current selves into our memories of the past and our anticipations of the future. As Velleman says, when he is remembering a five-year-old blowing out candles at his birthday party, he “conflate[s] [his] remembering self with the self of the experience remembered,” so he ends up thinking that he himself was back there blowing out those candles.

Velleman then goes on to argue that the idea of an enduring self is intimately connected to the idea of the passage of time, and that to give up the former would also lead to giving up the latter. (He argues for their connection in part by considering various views about the nature of time, including eternalism and presentism, eventually giving a sustained argument against presentism.) The conclusions that Parfit comes to about an enduring self are not radical enough precisely because Parfit still attempts to hold onto the passage of time. Velleman argues that we should reject them both.

Finally, Velleman makes some intriguing suggestions about how rejecting these two ideas can help us to cope with suffering, the prospect of death, and general anxieties of life, just as Buddhists have always thought.

**Daniel Dennett, “Where Am I?”**

This article is an extended thought experiment, meant to help the reader think philosophically about issues involving personal identity and survival. Here’s the basic idea.

Dennett describes an important job for which he was recruited by the Pentagon. They want him to go underground to retrieve a missile that got lodged under Tulsa, Oklahoma, and because the missile is extremely radioactive, they want him to leave his brain behind. (This is a fictional story, by the way.) He agrees and undergoes an operation to have his brain removed, placed in a vat, and connected via radio links to his body. He wakes up, goes to visit his brain floating in the vat, and he asks, “Where am I?”

It’s a tough question, given that his brain is in the vat at which he is now looking and his body is standing outside the vat looking in at his brain. He begins to think about the question philosophically, and gives names to his brain and his body to keep things straight. He names his bodiless brain “Yorick” and his brainless body “Hamlet.” He knows where Yorick is—in the vat. And he knows where Hamlet is—outside the vat. But the question is, Where’s Dennett? He runs through three possibilities: (1) Where Hamlet goes, there goes Dennett; (2) Where Yorick goes, there goes Dennett; and (3) Dennett is wherever he thinks he is. The problem is that each of these choices seems wrong.

Against option 1, Dennett appeals to the fact that it seems like he would be able to survive a simple brain transplant operation, in which his brain is placed in a different body and his old body is destroyed. If he would survive such an operation, then he can’t be where Hamlet is, as Hamlet is just his old body. Against option 2, Dennett points out that if he were to commit a crime and the police were to lock up Yorick, Dennett would still feel like a free man. Locking up Yorick doesn’t accomplish the goal of locking up Dennett, because Hamlet would still be free to roam the world. So, Dennett can’t be Yorick either. But option 3 is equally unsatisfying, because there are other circumstances in which people can temporarily change their point of view without thereby changing their location (for example, in a factory where someone is using a robotic arm or in a movie theater).

He fails to settle the question at this point, so he sets off for his underground adventure. As he is down by the missile, his body begins to lose contact with his brain back in Tulsa. Slowly he loses his five senses until finally he can neither feel nor see nor hear nor taste nor smell anything. Yet he is still thinking to himself—he decides that he is now back in the vat in Tulsa, having completely lost touch with the body that is still underground. He cannot communicate with anyone and begins to panic. But his team of scientists tranquilizes him for a while, and then he wakes up to hear them talking to him. He doesn’t have ears, of course, but they have plugged a microphone directly into his auditory nerve and are reassuring him that they are going to try to “re-embody” him as soon as they can.

Dennett gets a new body, which he calls “Fortinbras,” and although he looks quite different in the mirror, he still feels like the same old Dennett. He goes back to the room where Yorick is to pay him a visit, but he finds out that Yorick is no longer in control of his new body. The scientists have created a computer program that exactly matches the inputs and outputs of Yorick so that Dennett doesn’t even know the difference. Dennett’s new body is now being controlled by this new “brain,” which he calls “Hubert.”

A switch allows Dennett to toggle between having Fortinbras controlled by Yorick and having Fortinbras controlled by Hubert. He can switch back and forth without even noticing the slightest difference. But a worry remains: What if someone were to take the “brain” that Dennett wasn’t using at the moment and attach it to a different body? There would certainly then be two people, but which would be Dennett? Or consider a case not so extreme: What if the technicians just disconnected Yorick from Fortinbras altogether. Would there be two people in this case, too—the one Fortinbras-Hubert pair and the other Yorick? Which would be Dennett?

Because of these disconcerting questions, Dennett decides that the lab should be completely locked up, and Dennett carries the only master switch with him, unmarked, so that he never knows whether Yorick or Hubert is controlling his body.

At the end of the article, Dennett decides to flip the switch, only to find out that something has gone wrong. The other brain—Hubert or Yorick, we don’t know—has gotten out of synch with what has been going on with Dennett for a while so that the other brain (person?) realized he wasn’t really controlling anything and hence was a sort of prisoner inside of Fortinbras. When Dennett flips the switch, this brain (which seems quite upset with its plight) is glad to have been let out, and he plans to fly back to Houston to get one of them a new body.

**Marya Schechtman, “Personhood and Personal Identity”**

In this paper, Marya Schechtman argues against the psychological-continuity view of personal identity. This view notoriously faces problems with circularity, and Schechtman argues that the solution contemporary theorists have devised to address this problem doesn’t work. Schechtman then diagnoses why such views cannot succeed in escaping the circularity problem, arguing that these theorists conflate two questions. Attention to the second of these questions, Schechtman holds, reveals an interesting philosophical project concerning personal identity, which theorists have overlooked. Schechtman concludes by gesturing at how we might approach that project, in light of her discussion.

Schechtman begins her paper by distinguishing two questions: the question of reidentification and the question of self-knowledge. Traditionally, she explains, the question of personal identity has been thought to concern the first of these two questions: it has to do, in other words, with which history one’s life is a continuation of. Contemporary theorists have generally focused either on bodily or psychological continuity. The latter kind of view is more popular, and this is the kind of view on which Schechtman focuses.

In Section I, Schechtman lays out Derek Parfit’s particular psychological continuity view, which takes it that psychological continuity is a matter of one’s being strongly psychologically connected a person through various mental states. Parfit’s view is representative of typical psychological-continuity views, and it reflects the reductionist project such views undertake. Such views, that is, aim to analyze the concept of personal identity without referencing that concept. Circularity is devastating to such a project. In section II, Schechtman turns to such a circularity problem for psychological continuity views. The problem arises from the possibility of delusional memories (or intentions, desires, etc.) The psychological continuity view needs to be able to distinguish between real and apparent memories, but the only way to do so seems to be to appeal to whose memory it is. And this means referring to personal identity. The solution that Parfit and others suggest to this problem is to distinguish memories from “quasi-memories.” I have an accurate quasi-memory if I seem to remember having some experience, someone did have that experience, and my apparent memory is causally dependent, in the right way, on that past experience.

In section III, Schechtman argues that this solution doesn’t work. Parfit focuses on the relation between my seemingly remembering something and what happened. But, Schechtman argues, he ignores the reference the content of our memories make to ourselves. If I have a quasi-memory that isn’t my own, I will either need to have so much of my psychology changed so that I can relate to it that I will be effectively delusional, or my relation to the quasi-memory will not be such that it can establish the sort of psychological connectedness on which the continuity view relies. Finally, Schechtman argues, in section IV, that the problem arises from a conflation of the question of reidentification and the question of self-knowledge, which has to do with one’s self-conception. The reductionist project reflects the first question, but the second sort of question motivates the psychological continuity view. This leads Schechtman to reimagine the philosophical project of accounting for personal identity.

**Agnieszka Jaworska, “Respecting the Margins of Agency: Alzheimer’s Patients and the Capacity to Value”**

In “Respecting the Margins of Agency: Alzheimer’s Patients and the Capacity to Value,” Agnieszka Jaworska addresses an important dilemma facing many people who care for people with Alzheimer’s disease: in making decisions about how to care for them, should we privilege their previously professed attitudes, or should we privilege the attitudes they currently profess. This dilemma becomes particularly pointed when the previously professed attitudes conflict with the currently professed ones. Jaworska argues that we shouldn’t necessarily privilege past attitudes over their present counterparts. This is because, Jaworska contends, Alzheimer’s patients, despite their diminishing capacities, are nevertheless capable of forming new critical interests, new *values*. And such values, Jaworska holds, carry special weight for our well-being and should be respected.

Jaworska makes these points in dialogue with Ronald Dworkin, who argues that we should only privilege Alzheimer’s patients past critical interests, their past judgments about what is good, when determining how to care for them. Dworkin’s argument hinges on two claims. First, he claims that Alzheimer’s patients are incapable of having a sense of their life as a whole, and, second, he claims that having a sense of one’s life as a whole is essential to being able to form new critical interests. Jaworska argues against this latter claim.

She begins by suggesting that we can understand critical interests as being one’s values. And then she goes on to argue that valuing does not require one have a sense of one’s life as a whole. Moreover, she argues, Alzheimer’s patients have the capacity to value.

To make these points, Jaworska provides an account of valuing’s key marks: (1) that one thinks one is correct in wanting what one values; (2) that achieving what one wants is tied up with one’s self-worth; and (3) that the importance of what one wants is, for oneself, independent of one’s own experience. She then goes on to adduce cases in which Alzheimer’s patients, who don’t have a sense of their lives as a whole, nevertheless exhibit these key marks. Jaworska also adduces evidence from neuroscience, arguing that, in Alzheimer’s patients, the parts of the brain associated with valuing deteriorate after the parts of the brain associated with having a sense of one’s life as a whole.

One’s critical interests—one’s *values*—are weighty when it comes to one’s well-being, and thus Jaworska holds that we should take seriously Alzheimer’s current critical interests, and not only their past critical interests, when making decisions concerning their welfare. Moreover, the capacity to value, which Jaworska argues that Alzheimer’s patients have, is the starting point of autonomy, and while Alzheimer’s patients may not be able to execute their values in practice, Jaworska holds, we should nevertheless respect them and, as caretakers, help them execute their values.

**Roderick M. Chisholm, “Human Freedom and the Self”**

In this article, Chisholm sets out what he takes to be the problem of human freedom and then develops his own response to the problem, which requires some fairly radical claims about the nature of the human person.

The problem of human freedom, according to Chisholm, is as follows: Humans are responsible agents, but this fact appears to conflict with both a deterministic view of human action and an indeterministic view of human action. So, it appears that we have no way to account for the fact that humans are responsible agents.

First, consider a deterministic view of human action. Chisholm thinks responsibility is incompatible with a view like this, because he thinks that, in the case of a man being responsible for firing a shot, responsibility requires that “there was a moment at which it was true, both that he could have fired the shot and also that he could have refrained from firing it.” But for this to be the case, the act cannot have been caused or determined by some event that was not under the man’s control.

Second, responsibility appears to be incompatible with an indeterministic view of human action. Chisholm says, “If the act—the firing of the shot—was not caused at all, if it was fortuitous or capricious, happening so to speak out of the blue, then, presumably, no one—and nothing—was responsible for the act.” What, then, can we say? Should we give up responsibility?

Fortunately, Chisholm thinks that there is a *via media* between deterministic and indeterministic accounts of human action. His suggestion is that we move to a picture of causation according to which, sometimes, humans can cause events (as opposed to events causing other events). Or, in Chisholm’s terminology, we should say that not all the causation in the world is *transeunt* causation (between two events), but that some is *immanent* causation (between an agent and an event). In the immanent causation picture, the human agent does cause the action (so it is not an indeterministic picture), but the agent could have refrained, as well (so it is not a deterministic picture).

Chisholm goes on to consider two objections to this picture. First, someone might object as follows: We know that actions are actually caused by the goings-on in a person’s brain, and because some agent may not even be aware that he has a brain, that person certainly doesn’t *do* anything to his brain. But if he doesn’t do anything to his brain, then why not just say that it was some event in his brain that caused his action?

The reply to this objection relies on a distinction between “making something, A, happen,” and “doing A.” Chisholm thinks that we can’t infer from the fact that an agent doesn’t do anything to his brain that he doesn’t also make something happen in his brain. Although he may not intend the goings-on in his brain, nevertheless, by causing his action, he makes those particular goings-on occur.

The second objection asks the following question: What is the difference between A’s just happening and the agent’s causing A to happen? In other words, what does immanent causation add to the picture?

Chisholm responds with two points. First, one thing that makes immanent causation different from A’s just happening is that in the first case, A was caused by the man, whereas in the second case, A had no cause at all. Chisholm acknowledges that this is not a particularly satisfying response, but his second point has more force. He points out that this problem is not unique to the proponent of immanent causation and can be raised even against someone who thinks all causation is between events. We could ask, for example: What is the difference between saying that B happened and then A happened and saying that B’s happening was the cause of A’s happening?

**David Hume, “Of Liberty and Necessity”**

In this classic selection, Hume treats the problem of reconciling liberty and necessity. Or, as we might cast the problem in contemporary terms, that of reconciling free will and determinism. Hume contends that the entire debate rests on mere terminological confusion, and that once we clear up the terminology, we will see that there is no tension between liberty and necessity at all.

Hume first argues that everyone shares the same idea of necessity, an idea that is derived from the constant conjunction we observe in the operations of nature. We see that certain causes are always followed by certain effects in the physical world, and the same lesson can be transferred to apply to humans. The same motives always produce the same actions across cultures and centuries. Everyone has always recognized this and, therefore, has agreed about the doctrine of necessity. Why, then, have people disagreed about liberty?

Hume explains that because people don’t “feel” any connection between their motives and their actions, they thus conclude that their actions must not be the result of any sort of necessity. But this, according to Hume, is a conclusion that is much too hasty. The real way to view liberty is merely as “a power of acting or not acting, according to the determinations of the will.” As Hume puts it, we have liberty just in case if we choose to remain at rest, we may, and if we choose to move, we may. Hume goes on to argue that liberty in this sense is not at all in conflict with necessity in the preceding sense.

There are a few other arguments that Hume considers that also lead him to the conclusion that liberty and necessity can indeed be reconciled.

**Harry G. Frankfurt, “Alternate Possibilities and Moral Responsibility”**

In this article, Frankfurt mounts an attack against a (once) widely held principle about moral responsibility, the principle of alternate possibilities (PAP). As Frankfurt formulates it, PAP runs as follows: A person is morally responsible for what he has done only if he could have done otherwise. Frankfurt thinks this principle is false, and his attack proceeds as follows.

Consider the case of Jones. Frankfurt describes the case as follows: “Jones decides for reasons of his own to do something; then someone threatens him with a very harsh penalty (so harsh that any reasonable person would submit to the threat) unless he does precisely that, and Jones does it.” Now, Frankfurt asks, is Jones morally responsible for what he has done? Our answer to this question is so far underdetermined by the description of the case. Each way we might fill out the details of the scenario will issue in a particular judgment.

First, say that the threat had no force on Jones whatsoever. (Frankfurt calls this man Jones1.) In this case, Jones1 is not coerced at all. He is morally responsible for what he has done, and he has alternative possibilities. This case matches up with PAP so far.

Second, say that the threat was so alarming to Jones that he completely forgot his initial intentions of performing the action, and he goes on to perform the action solely because of the threat. (Frankfurt calls this man Jones2.) In this case, Jones2 is coerced. He is not morally responsible for what he has done, and he has no alternative possibilities. Again, this case matches our intuitions about PAP.

Now comes the tough case. In the case of Jones3, the threat had some effect on Jones3, but not very much. More specifically, the threat affects Jones3 as follows: In the actual course of events, Jones3 is not alarmed by the threat, and he continues to perform the action for his own reasons. However, had he not already made the decision to perform the action, he would have been so alarmed by the threat that he would have submitted. In this case, Frankfurt maintains, Jones3 is morally responsible for what he has done, and he does *not* have alternative possibilities.

These considerations lead Frankfurt to conclude that the case of Jones3 is a straightforward counterexample to PAP. But he anticipates an objection that runs as follows: There is a sense in which Jones3 cannot do otherwise than perform the action, but this sense doesn’t tell us anything about whether he has alternative possibilities. After all, just because he has been threatened doesn’t mean that he *can’t* do otherwise—it just means that the consequences of his doing otherwise will be bad.

But Frankfurt thinks he can avoid the objection by reformulating the case of Jones. Consider now Jones4. Let Black be the person who would issue the threat to Jones4, but instead of having Black actually threaten Jones, let Black be a mere *counterfactual intervener*. That is, let Black monitor Jones4’s brain (or in some other way predict what Jones4 will do), and only step in if Jones4 is about to decide to do whatever Black does not want him to do. As the case actually plays out, Jones4 does what Black wanted him to do, so Black never has to step in. But Black was nevertheless poised so that it is nevertheless true that Jones4 cannot do otherwise. In this case, Frankfurt maintains, Jones4 is morally responsible even though he has no alternative possibilities.

**John Martin Fischer, “Responsiveness and Moral Responsibility”**

In “Responsiveness and Moral Responsibility,” John Martin Fischer sketches an account of moral responsibility according to which an agent is morally responsible only if she is capable of responding to reasons. He begins by considering a famous case, viz., “Frankfurt-type” cases. In these cases, an agent putatively lacks the ability to do otherwise (i.e., alternative possibilities) and yet, intuitively, she is morally responsible for her action. These cases have led Fischer to conclude that moral responsibility doesn’t depend on an alternative sequence—i.e., how things could have gone—but instead, an agent’s being morally responsible depends on the actual sequence—that is, how things actually go.

According to Fischer’s actual-sequence theory of moral responsibility, an agent is morally responsible for her actions only if, in the actual sequence, she is reasons-responsive. Fischer distinguishes between two forms of reasons-responsiveness: strong and weak. But he concludes that only the weak form of reasons-responsiveness is necessary for moral responsibility. Fischer further explicates weak reasons-responsiveness by comparing it to accounts of knowledge according to which an agent knows that *p* only if her belief is sensitive to the truth in a particular kind of way. Similarly, an agent is morally responsible for her action only if the mechanism that her action issues from is sensitive to the reasons for action in a particular kind of way.

Fischer concludes by exploring what he calls “semicompatibilism.” Semicompatibilism takes seriously the argument that the freedom to do otherwise is incompatible with causal determinism. But it also takes seriously the insights that Fischer has developed in this paper, viz., that moral responsibility is an actual sequence property and therefore does not require the freedom to do otherwise. Thus, according to Fischer, the truth of determinism would not rule out moral responsibility because determinism is no threat to weak reasons-responsiveness.

**Harry G. Frankfurt, “Freedom of the Will and the Concept of a Person”**

In this important article, Harry Frankfurt gives some intriguing reflections about persons, the will, and freedom. He develops a subtle account of the hierarchical structure of a person’s will and relates that account to various problems concerning freedom and moral responsibility.

Frankfurt begins by asking, “What is a person?” Although many answers suggest themselves, Frankfurt thinks there is one specific characteristic that can serve to distinguish persons from nonpersons: the hierarchical structure of a person’s will. Persons, as Frankfurt understands them, are beings that have second-order desires. That is, persons are capable not only of wanting to do certain things but also of wanting to want to do certain things.

But it is not enough to be a person merely to have second-order desires, according to Frankfurt. You must also have some second-order volitions. A second-order volition is a second-order desire that a particular first-order desire will motivate the agent to action. All and only beings with second-order volitions, in this sense, are persons.

To illustrate the idea of a second-order volition, Frankfurt gives the case of the willing addict and the case of the unwilling addict. Both addicts have first-order desires to take the drug that eventually motivate them to, in fact, take the drug. But whereas the willing addict wants those desires to motivate him, the unwilling addict does not want those desires to motivate him.

Moreover, all and only beings with second-order volitions have free will, according to Frankfurt. To be free with respect to an action is to be able to do what you want to do. But to be free with respect to your will is to be able to will what you want to will. Only beings with second-order volitions can have this further characteristic.

Frankfurt goes on to apply these notions to various cases and draws some interesting conclusions about moral responsibility and determinism.

**Gary Watson, “Free Agency”**

A person is generally thought to be free when she is able to do or get what she wants. This conception of freedom faces a problem, though. Namely, we are often inclined to say that a person isn’t free when she is not in control of her own actions, despite the fact that her behavior is intentional. For example, consider compulsive agents—like kleptomaniacs or people with phobias. In “Free Agency,” Gary Watson offers an account of agency that is meant to explain why such agents are unfree, without compromising our ordinary thoughts about freedom. Moreover, this account shows that whether an action is free or unfree has nothing to do with the truth or falsity of determinism.

Watson distinguishes between two classical ways of conceiving of practical reasoning. On the Humean model, Reason is distinguished from Passion, and Reason serves an instrumental purpose, whereas Passion is a source of motivation. Alternatively, on a Platonic model, Reason is distinguished from Appetite, and both serve as sources of motivation. Watson takes himself to be offering a Platonic model.

For Watson, we have two “systems.” Our valuational system represents our practical standpoint—our agential identity. This system issues in judgments about what the best thing, overall, to do is. It overlaps partly with our motivational system, which represents the set of considerations that move us. Particularly, our actions are free when these systems are in harmony. However, we have different sources of motivation, not all of them part of our valuational system. And when we are moved by these other sources of motivations—appetitive desires, emotions, desires we’ve acquired from acculturation—our actions are unfree.

In the last part of his paper, Watson offers a criticism of Harry Frankfurt’s famous article, “Freedom of the Will and the Concept of a Person.” For Frankfurt, free action is a matter of forming second-order volitions—desires about our first-order desires’ efficacy. Watson notes, though, that a second-order volition is just a desire, like any first-order desire. It is unclear, then, why a second-order volition has any authority. Why, that is, are we identified with our second-order volitions and not our first-order desires? There is no principled difference between the two. Watson suggests that valuing serves the role better.

**Christian Helmut Wenzel, “Free Will and Zhuangzi: An Introduction”**

In this commissioned piece, Christian Wenzel focuses on the ancient Chinese text, the *Zhuangzi*, and how it relates to contemporary themes in free will. Wenzel argues that the *Zhuangzi* is both surprising and illuminating when contrasted with contemporary work.

Wenzel focuses on five notions from the contemporary literature: (1) the freedom to do otherwise; (2) authorship; (3) being responsive to reasons; (4) control; and (5) moral responsibility.

Interestingly, (1) hardly shows up, if at all, in the *Zhuangzi*, while the other four appear in some ways, but often in different contexts than is typical of the contemporary literature. As Wenzel shows, a central theme of the *Zhuangzi* is the idea of spontaneity—that instead of engaging in reflective deliberation, say, to express one’s agency, one ought to act skillfully but without much thought. A central example concerns that of an experienced woodworker, one that is able to carve a piece of wood in a masterful way and without any thought. Similarly, proclaims the *Zhuangzi¸* sowe ought to act in our day to day lives.

With this in mind, the notions of (2), (3), (4), and (5) look quite different than in the contemporary literature. The notion of authorship, Wenzel claims, effectively disappears; the notion of being responsive to reasons is more appropriately understood as being responsive to situations; the notion of control is radically changed from one of an active sense to that of a passive one; and the notion of moral responsibility involves trusting what is natural or intuitive rather than what one can arrive at through reason and deliberation. In the end, Wenzel argues that these suggestions from the *Zhuangzi* are often insightful, although they aren’t without their problems.