Altruism and the Separateness of Persons

The conventional view of morality is that there are reasons of an impersonal kind, or reasons for promoting other peoples' well-being. I shall call these reasons for altruism. Does reflection on the nature of personal identity lend any additional support to the conviction that there are reasons for altruism? A growing number of contemporary philosophers claim that it does. Their common theme is that the boundaries that separate persons are less substantial than is commonly supposed. This revision to our theory of personal identity has consequences for practical reason. One consequence is that it strengthens the case for altruism.

I shall call arguments of this general kind arguments for altruism from the unity of persons—or simply the unity argument for altruism. Versions of this argument have been put forward by Derek Parfit, Christine Korsgaard, David Brink, and Carol Rovane.¹ The strategies they develop in support of the unity of persons argument have some variations (as will emerge below when I distinguish between psychological and rational altruism, as well as content-conservation and historical versions of the agency theory). But there is a common premise that underpins their views. We can call the relation that provides the identity conditions for persons the unity relation. In order for the argument for altruism from the unity of persons to succeed, it must be true that the same kind of relation that unifies people over time normally extends between bodies separated in space. Putting the point another way, we bear the same kind of relation to our future self—normally the future occupant of the body we now occupy—as we do towards individuals in other bodies. Since the unity relation obtains both between bodies that are separated by time and be-


© Copyright 2001 by Social Theory and Practice, Vol. 27, No. 3 (July 2001)

361
tween those that are separated by space, it can truthfully be said that the separation of persons is not deep. These observations about personal identity support varieties of impersonal concern. Or so it is claimed.

In order for this intriguing argument to succeed, however, it will be apparent that the unity relation for persons must be characterized in a particular way. The basic restriction is that this relation must be of a kind that normally extends across bodies separated by space. Otherwise, there will normally be a significant difference between the relationship we bear towards our future self and our relationship to minds located in other bodies. This asymmetry would, in turn, undermine the thesis that the case for altruism derives special or independent support from revisions to the theory of personal identity.

What this means is that proponents of the unity argument for altruism must reject certain venerable proposals about the nature of personal identity. They must resist the suggestion that physical continuity is central to personal identity, since we do not normally share a body or brain with others. They must similarly reject the neo-Lockean thesis that personal identity comprises continuity of consciousness. For our access to the experiences hosted by the body we occupy is normally very different from our access to experiences hosted in other bodies. In short, the unity relation for persons cannot comprise either physical or phenomenological continuity. It scarcely bears mention that if persons are individual substances rather than relations, the argument for altruism from the unity of persons does not get to the starting blocks.

Proponents of the unity argument for altruism respond by defending a rival account of personal identity. They analyze personal identity in terms of the continuity of agency—roughly, as continuities in plans, values, and related practical attitudes. I emphasize that this is a rough characterization. As will emerge, there is some disagreement about the nature of the agency relation. But these differences are relatively minor compared to the points on which proponents of the unity argument for altruism must be in agreement. To repeat, unless persons are some kind of relation that excludes physical and phenomenological continuity, then the separateness of persons remains deep. It is perhaps not as deep as it might be if each individual were an independent soul. But unless physical continuity and the phenomenological continuity it normally sustains are excluded from the unity relation, then a reexamination of personal identity supplies no independent grounds for altruistic concern.

My central contention in this essay is that personal identity cannot plausibly be analyzed in terms of the continuity of agency. The continuity of agency is not necessary for personal identity. Nor is it sufficient. Or so I shall argue. But before I turn to those claims, I will provide a fuller exposition of the argument for altruism from the unity of persons.
1. The Connection Between the Theory of Personal Identity and Altruism

The thesis we are investigating holds that altruism receives independent support from reflection on the nature of personal identity. It will be useful to distinguish some ways in which that support is thought to come about.

We can get started by attending to some matters of terminology. I will define altruism as a non-derivative concern for other peoples’ well-being. S may knowingly advance T’s well-being on the grounds that this will ultimately benefit S. In this circumstance, T’s well-being has instrumental value for S. For S regards T as a means to ends that do not essentially include a concern for T’s welfare. If S could achieve his goals without T, or through T’s misery, that would make no difference to S. An instrumental concern for others does not qualify someone as an altruist.

The purest cases of altruism feature a concern for another individual qua member of the human species or qua rational being. In these cases, we value others not on account of what they can do for us, but in virtue of their humanity. But altruism may also come in less Kantian flavors. It can obtain between the members of less inclusive associations: between friends, family members, and compatriots. We often harbor a concern for these individuals that goes beyond their status as means to our ends. We may regard their happiness as being one of our ends. Someone’s actions qualify as altruistic whenever they issue from a non-derivative concern for others. But it needn’t be the case that altruists aim to be evenhanded in their benefactions.

We must next distinguish between psychological altruism and reasons altruism. Psychological altruism is a thesis about motivating concerns. It purports to describe an individual’s passions or conative states. Psychological altruism is a true description of S, if S is motivated to promote other peoples’ interests out of non-instrumental considerations. Reasons altruism, by contrast, is a thesis about the kinds of aims an individual should or ought to have. Reasons altruism presents a norm for practical reasoning. That norm expresses the imperative that individuals ought to advance the well-being of others on non-instrumental grounds.

Now there are versions of reasons altruism for which the statement “S has a reason to benefit T” is true only if S is motivated to benefit T. Let us call this view subjectivist or internal reasons altruism. According to internal reasons altruism, S has a non-derivative reason to benefit others only if some (possibly counterfactual) version of S—normally a fully informed or epistemically idealized version of that agent—desires to benefit others.² An implication of internal reasons altruism is that all

²For an influential statement of subjective reasons for altruism see R.B. Brandt, A
agents have reasons for altruism only if psychological altruism is universally true. By contrast, according to objectivist or external reasons altruism, the grounds for altruism are independent from an agent’s desires to benefit anyone. Facts about people’s motives don’t determine whether they have reason to benefit others.3

These distinctions are indispensable for coming to terms with the argument for altruism from the unity of persons. As will emerge, some proponents of the argument make the comparatively modest claim that an examination of personal identity supports psychological altruism. Their claim is that someone could sensibly acquire altruistic motivations as a result of coming to understand the truth about persons. But other proponents of the unity argument are more ambitious. Their focus is not primarily psychological. They assert rather that the facts about personal identity demonstrate that agents normally have reasons for altruism. Crucially, these are reasons in the objective or external sense. Reflections on personal identity show that an agent’s reasons for altruism are independent from her actual or counterfactual desires. Putting this point another way, reasons altruism is a logical rather than a merely psychological thesis. The reasons for altruism can be inferred from an analysis of the concepts of personal identity and practical reason, in conjunction with certain facts that bear no essential connection to what motivates a given agent.

The difference between psychological and reasons altruism will emerge more clearly as we briefly survey some recent presentations of the argument for altruism from the unity of persons.

Parfit’s View: Parfit argues that persons are psychological relations (of a kind that will emerge more clearly in section 2). He claims further that the unity relation for persons can obtain by degrees. The common notion that a person is either alive or dead does not withstand scrutiny. It is also a matter of empirical fact, according to Parfit, that the psychological relations constitutive of persons normally diminish over time—even if the brain hosting those relations suffers no extensive amnesia, loss of affect, or other mental disorder that can raise special worries about a loss of psychological continuity. The unity relation for persons diminishes over time even when the mind is sustained by a normally functioning brain.4

---

Theory of the Good and the Right (Oxford: Clarendon Press, 1979). Brandt claims that it would be rational to endorse a utilitarian morality. For he identifies rational desires as desires that survive cognitive psychotherapy, and he predicts that altruistic desires will withstand this exposure to the facts and logic (p. 126).


4I will use the term “survival” to designate those occasions when the relations con-
Parfit maintains that most people believe they are Cartesian souls. They believe, in Parfit’s parlance, that there is some “deep fact” of that kind that makes them persons. The discovery that we are relations, rather than an immaterial substance, should have an important psychological impact. It should erode our sense that there is something especially significant about our body and its limits. For bodies can no longer be thought to acquire an importance based on the notion that each body plays host to a soul for the duration of the body’s existence.

This discussion of personal identity spills over into Parfit’s revised account of prudential reason. Prudence directs us to make sacrifices now for the benefit of our future self. These sacrifices are regulated by a discount rate, a rate that determines the amount at which current benefits should be forgone for the sake of our future well-being. It is usually assumed that our future self is coextensive with the future occupant of our body (more or less). Parfit maintains, however, that we should heavily discount the well-being of the future occupant of the body we now occupy. This is on the grounds that the future occupants of that body border on being separate persons with the passage of time, and prudence does not require us to sacrifice our current well-being for the benefit of others. Prudence is therefore less demanding than is commonly supposed.  

But then does practical reason have nothing to say about what we would ordinarily describe as the domain of intrapersonal relations? No, because according to Parfit, the scope of morality expands to fill the breach formerly occupied by prudence. Morality regulates the relationships between the occupants of a single body separated by large chunks of time. For these relationships are no different from the interpersonal relationships that are commonly thought to fall under morality’s jurisdiction. Parfit's remarks on prudence help to make concrete his suggest-
tion that the limits of the body are of no special significance from the standpoint of practical reason.

This discussion conceivably lends support to altruism in several related ways. First, if Parfit is right, then learning the truth about persons—that they are not grounded in some “deep fact”—can stun our ordinary form of personal concern. Second, Parfit’s discussion arms him with a reply to someone who opposes altruism on the specific grounds that the boundaries of the body are terribly significant from the standpoint of practical reason. That argument is mistaken. For the lines of jurisdiction between prudential and moral reasons don’t match the limits of the body—at least if we accept Parfit’s claim that the people regularly perish in normally functioning bodies.

Let us suppose that this characterization of Parfit is accurate. (If it is incorrect, that should not make a difference to the grounds I later give for rejecting the argument for altruism from the unity of persons.) If his arguments are plausible, they lend support to psychological altruism. But they don’t seem intended to support reasons altruism. To be perfectly clear, Parfit indeed believes there are moral or impartial reasons. And he also believes that reflections on the nature of personal identity have consequences for the content of morality. For example, distributive justice should not, in his view, compensate people for misfortunes “they” have suffered in the past, since the present and past occupants of a single body may be separate persons. But what Parfit does not claim is that a revised understanding of personal identity lends direct support to the thesis that there are external reasons for altruism. Nor does he claim that altruistic reasons are a species of prudential or self-interested reasons. The grounds for impersonal reasons are independent from revisions in the theory of personal identity.

**Brink’s View:** David Brink advances the more ambitious claim that a revised conception of personal identity lends independent support to the thesis that altruism is rational. Moreover, the reasons for altruism are external in the sense that their warrant is unconnected to an agent’s desires. Someone who failed to acknowledge the authority of these reasons

---

6This is a big “if,” since the truth about persons may induce nihilism—the view that nothing matters—or a variety of other psychological reactions. In this vein, Korsgaard describes at length why we should be skeptical of Parfit’s psychological claims. According to her, our self-concern is not grounded in the belief that we possess a soul, but rather in pragmatic pressures of a general kind: chiefly, the necessity for acting. But she agrees with Parfit’s metaphysical thesis that persons are psychological relations. She also seems to agree that reflection on the truth about persons lends support to altruism. I use the term “seems” advisedly, however, since it is unclear what kind of support she reckons these reflections provide for altruism. My reservations will emerge more clearly in section 2.
would normally be guilty of drawing arbitrary distinctions within the domain of prudential reason.

Brink’s argument goes as follows. Like Parfit, he maintains that personal identity comprises the right kind of psychological relations. And like Parfit he claims that prudential reasons are external. Brink emphasizes, however, that the relations that unify persons over time also obtain between bodies that are separated in space. The psychological relations that are constitutive of personal identity normally extend outside of the body—albeit in a diminished way. Here Brink seems to be drawing attention to a consequence of Parfit’s theory of personal identity. (Whether Parfit would endorse that consequence or anticipates it does not concern me here.) Brink then advances the novel thesis that prudence requires us to promote the well-being of those entities to whom we are connected by the unity relation.

Let’s consider this proposal in greater detail. Here are some characteristic passages from Brink:

If ... the separateness or diversity of persons is not so fundamental... Insofar as distinct individuals are psychologically connected and continuous, each can and should view the other as one who extends her own interests in the same sort of way that her own future self extends her own interests. (142)

This means that each should regard the good of those to whom she stands in such relationships as a constituent part of her overall good, just as she should regard the good of her own future self as a constituent part of her overall good. (143)

Interpersonal connections and continuity can be found among intimates who interact on a regular basis and help shape each other’s mental life; in such relationships, the experiences, beliefs, desires, ideals, and actions of each depend in significant part upon those of the others. (141)

I shall return to issues surrounding the unity relation—the psychological connectedness and continuity that Brink refers to in the first passage—in section 2. For now I want to concentrate on the conceptual connection between personal identity and altruism in Brink’s account. There is no doubt that we normally care a great deal for the happiness of people with whom we “interact on a regular basis.” It is a matter for personal concern whether our friends, lovers, or children are happy. It is normally correct to say therefore that agents are motivated to advance the well-being of their intimates. If there are such things as internal reasons, then it’s equally correct that we possess internal reasons to benefit our intimates. But Brink is claiming something different. He states that “each should

---

regard the good of those to whom she stands in such relationships as a constituent part of her overall good." The operant term in this passage is the imperative "should." Even those individuals who lack personal concern for their intimates can have reason to promote the latter's well-being. The grounds for these reasons are independent from their desires and related subjective states. The grounds relate to whether the unity relation obtains between themselves and other individuals. In short, Brink's thesis is that some prudential reasons are external reasons, and reasons for altruism inherit this external status since they are a species of prudential or self-interested reason.

**Korsgaard's View:** There is some question about how much of Brink's position was anticipated by Christine Korsgaard. In an early critique of Parfit she wrote:

This is not to say that the considerations against deep personal separateness that Parfit and I both endorse have no consequences for the standard philosophical model of rationality. I have suggested that agents come in different sizes, and that the human body is merely the basic one. If we grant that the unity of agency is a reason for future concern, then we should grant that I also have reasons to care for the future of larger agencies of which I am a part. Just as I have a personal concern for my physical future, I may have a personal concern for the future of my family, the organization for which I work, a project in which I have been active, or the state of which I am a citizen. The territory of practical reasons is not split into two domains—self-interested rationality concerned with the occupant of this particular body on the one hand, and reasons of impartial morality on the other. Instead, the personal concern which begins with one's life in a particular body finds its place in ever-widening spheres of agency and enterprise, developing finally into a personal concern for the impersonal—an concern, that is to say, for the fate of one's fellow creatures, considered merely as such.  

This passage occurs in the context of an essay in which Korsgaard explains how Parfit's account of personal identity differs from the "Kantian" alternative that she favors. I'll say more about those differences later. But here Korsgaard signals her agreement with Parfit's thesis that persons are psychological relations rather than substances. She goes on to claim that this revised conception of persons should have consequences for practical reason. It should induce the acquisition of non-derivative and personal concern for the well-being of others, or what I am calling altruism. The question is, what foundation for this concern does she have in mind?

Korsgaard is certainly claiming that reflections on the nature of personal identity promote psychological altruism. We should acquire a concern for others that we normally extend towards our "physical future"—that is to say, the future occupant of the body we occupy now. But she

---

8Korsgaard, p. 383.
may also be advancing the more ambitious thesis familiar from the dis-
cussion of Brink, namely, that reflection on the nature of personal iden-
tity supports the conclusion that there are external reasons for altruism of
a prudential kind.

This second reading is supported by Korsgaard’s claim to have
reached an insight into the shortcomings of the “standard model of philo-
sophical rationality.” It is also suggested by her related observation that
splitting “the territory of practical reason” into the narrowly self-
interested and the impartial involves a mistaken dichotomy. Both these
remarks suggest that some deep revision to the theory of practical reason
is in order.⁹ On this reading, Korsgaard and Brink effectively agree on
the logical connections between a suitably revised theory of personal
identity and the reasons for altruism. If there is any disagreement in their
views, it pertains to the kinds of psychological continuity that comprise
the unity relation for persons.

I must admit that I am uncertain which interpretation of Korsgaard is
correct. She may simply be advancing an argument for psychological
altruism. Yet if that is all she is claiming, then it is difficult to understand
why her position calls for anything so dramatic as revisions to “the stan-
dard philosophical model of rationality.” She does not say what she con-
siders the standard model to be. But the internal theory of reasons can
easily accommodate her insight that agents have special concern for
friends, family, or their state. On the internal conception of a reason, there
is no dichotomy between egoistic reasons on the one hand, and im-
partial reasons on the other.

Korsgaard’s position could only pose a challenge to a standard con-
ception of rationality if she is claiming that the reasons for altruism are
external reasons. For the notion that there are reasons of an external kind
to benefit others is indeed controversial. If this is the right way to inter-
pret Korsgaard, then she and Brink hold similar views (although they
may disagree about the exact nature of the agency relation).

2. The Continuity of Agency

The question we have been considering pertains to the connection be-
tween revisions to the theory of personal identity and practical reason. I
now proceed to investigate the proposals about personal identity favored

⁹In this connection, Carol Rovane endorses the view that individuals connected by
the agency relation are members of group persons. But she distinguishes her position
from Korsgaard’s, on the grounds that the latter insists that “human-size persons are basic
agents.” Rovane maintains that the boundaries of the human body merit no special sig-
nificance or priority (The Bounds of Agency, p. 161, n. 13).
by advocates of the argument for altruism from the unity of persons.

Proponents of the unity argument are all agreed that persons are not Cartesian souls or substances. They are relations of a certain kind. But why should this discovery about the nature of persons render altruism more attractive? This question arises because there are many alternatives to the substance view, which do little to erode the common notion that the separation between persons is deep. If personal identity involves physical continuity, then this particular argument for altruism seems misguided. For on a physical theory there is a genuine difference—a difference salient within the theory of personal identity—in the relationship we bear towards the future occupant of the body we now occupy and the occupants of other bodies. Similarly, if the unity relation for persons comprises a phenomenological unity, this would again lend no support to altruism. We can have propositional knowledge that another person has an experience, believing that they are in a certain state of mind. But we lack the access to these experiences that we seem to possess for past inhabitants of the body we now occupy. The difference, roughly, is like the difference between reading a detailed autobiography of someone else's life and recalling how the events that filled our past felt when they were taking place. We likewise anticipate the future experiences hosted in the body we now occupy in ways in which we do not anticipate experiences hosted in bodies from which we are separated in space. If personal identity comprises phenomenological unity over time, then we should once again conclude that the separation between persons is deep.

10The phenomenological properties associated with experience memory and the question of whether one person can have someone else’s experience memories have most recently been considered in helpful ways by David Velleman and Arthur Collins. Distinguishing between what he calls personal and factual memories, Collins claims that the former is marked by an ability on the part of the person who has them to elaborate indefinitely on the state of affairs being recalled. In his example, they do not simply recall the proposition that “Toscanini conducted on evening X” but can furnish details of the performance in response to questioning. See his “Personal Identity and the Coherence of Q-memory,” The Philosophical Quarterly 47 (1997): 73-80, pp. 75, 79. In a subtle discussion, Velleman provides a detailed analysis of the phenomenology of experience memory. See his “Self to Self,” The Philosophical Review 105 (1996): 39-76. Velleman defends the possibility of person A having person B’s memories, while Collins is critical of quasi-experience memories.

11Raymond Martin maintains that the continuity of experience—rather than the continuity of value or agency—is central to survival. In a fashion reminiscent of the argument for morality from the unity of persons, he furthermore claims that people can rationally anticipate having other peoples’ experience: “the anticipation of having someone else’s experiences requires that the anticipator identify with that person, not in the same sense that the anticipator regards the other person as the same person as himself, but only in that many of the anticipator’s dispositions, particularly his affective ones, that normally take only himself as an object of concern now take the other person as an object of concern” (p. 319). See his “Having The Experience: The Next Best Thing To Being There.”
Altruism and the Separateness of Persons

Proponents of the argument for altruism from the unity of persons deal with this problem by defending a very particular account of the unity relation. I shall dub this view the agency theory of personal identity. It emphasizes that continuity in plans, values, and related practical attitudes are central to personal identity. There may be some genuine differences in the kinds of psychological relation favored by those who argue for altruism from a revised conception of persons. Indeed, some may bridle at the label “agency theory of personal identity.” But labels are not important. For the parameters of the debate are clear enough. If the argument for altruism from the unity of persons is to have any hope of success, then neither physical nor phenomenological continuity can be the relations that matter for personal identity. For both relations ground a genuine difference between our future self and what are conventionally thought to be other individuals. The adoption of the agency theory of personal identity does not guarantee success for the altruism argument, since that argument also depends on the controversial premise that prudential reasons are external (at least the version of the argument that concludes that there are external reasons for altruism). But an agency theory will need to be adopted if the argument for altruism is to have any hope of success.

Let us investigate agency theories of personal identity in greater detail. Parfit describes a member of the Russian nobility who, as young man, elicits a promise from his wife that she will help manage his estate for high-minded social purposes. As his idealism wanes he requests control over his assets with less noble aims in mind. He urges her to forget the vow she had earlier made to him. Parfit claims that the aristocrat’s wife should deny this request. This is on the grounds that the man is a numerically different person from the individual to whom she made her promise, in conjunction with the assumption that promises should only be broken with permission from the promise bearer.\textsuperscript{12}

On the basis of this example, I believe that Parfit subscribes to a version of what I am calling the agency theory of personal identity. For in Parfit’s description of the case there is no suggestion that the aristocrat

\textit{Philosophical Studies} 70 (1993): 305-21, p. 312. Thus, there is, in Martin's view, a real difference in the type of access we have to the experiences of the future occupant of our body, and the occupants of other bodies. But Martin goes on to claim that we should nonetheless anticipate the experiences of others on the grounds that these anticipations have morally desirable consequences (pp. 317-18). Whatever merits this argument has, it cannot lend independent support to the conclusion that there are reasons for altruism. For instead of arguing from a revised understanding of personal identity to the novel conclusion that we should be altruists, it argues from a prior commitment to altruism to novel revisions in the theory of personal identity.

has swapped bodies, or, alternatively, has suffered a debilitating form of amnesia. Presumably, he recalls the salient episodes in his gradual transformation from idealistic youth to blasé aristocrat. He does not remember those episodes, moreover, in the way that a biographer might record them. He recollects the relevant experiences: the letdown of discovering that there is less to agrarian socialism than he once thought, creeping despair over the loss of his ideals, and so forth. Yet despite these experience memories, and the phenomenological unity that they help comprise, Parfit claims that he is not the same individual. The only explanation for this assessment is that Parfit must be tacitly assuming that having qualitatively similar values over time is necessary for being the same person. This is a version of the agency theory of personal identity.\textsuperscript{13}

Parfit’s analysis suggests a content-conservation version of the agency theory. To determine whether two person stages are phases of the same individual we make diachronic comparisons between the contents of their practical objectives. They are phases of the same person if there is sufficient overlap in the practical aims held by those person-stages.

This characterization of the unity relation for persons is obviously a nonstarter for anyone who believes that physical and/or phenomenological continuity suffices for personal identity. Yet even some agency theorists could have trouble with Parfit’s conclusion. The gist of their reservations is that the psychological causes of someone’s practical outlook makes a significant difference. Consider two possibilities. The first is that the aristocrat abandoned his youthful ideals because he was brainwashed, hypnotized, or surgically intervened with. At the opposite end of the spectrum, he initiated the transformation in his practical outlook. He elected, for example, to attend University, fully cognizant of the fact that many students change their minds about politics in University; he intentionally placed his socialist ideals at risk by attending seminars organized by the Adam Smith Society, or what have you. Some philosophers embrace what I shall call a causal conception of the agency theory. Whether someone is the same person over time is not discovered through a diachronic comparison of the brute contents of the relevant practical attitudes. Rather, identity is conserved if the evolution in those attitudes has the right kind of psychological cause.

For example, Korsgaard defends what she terms a Kantian account of personal identity. In this vein, she writes that:

\textbf{we can distinguish beliefs and desires that continue merely because, having been acquired in childhood, they remain unexamined, from beliefs and desires that continue be-}

\footnote{\textsuperscript{13}Parfit’s official view, however, is that there is an irreducible plurality in the kinds of psychological events that comprise the connection and continuity relations he uses for defining personal identity (p. 205).}
cause you have arrived at, been convinced of, decided on, or endorsed them. In an account of personal identity that emphasizes agency or authorship, the latter kind of connection will be regarded as much less boring than the former. This is because beliefs and desires you have actively arrived at are more truly your own than those which have simply arisen in you (or happen to inhere in a metaphysical entity that is you).  

According to Korsgaard, it is the psychological processes that generate our attitudes—what she here calls “authorship”—rather than agreement in attitudes over time that matter for personal identity. This is an instance of what I am calling a causal conception of the continuity of agency.

Many questions remain outstanding from this brief survey. Exactly what kinds of psychological processes distinguish bona fide authorship from discontinuous agency? Are differences between content-comparing and causal versions of the agency theory genuine differences? \(^{16}\) We should resist the temptation to pursue these topics. They ultimately don’t matter for an overall assessment of the argument for altruism from the unity of

\(^{14}\)Korsgaard, p. 379.

\(^{15}\)Stephen White provides another variation of the causal agency theory of personal identity. The basic consideration that determines whether someone has survived, he claims, is a form of control. An individual who lacks control over the intentions of her future self or selves—because those future goals have been generated through hypnotic suggestion for example—would fail to have survived. White’s main target is Neo-Lockean theories of personal identity. (He emphasizes that it is the anticipation of a particular kind of control rather than the anticipation of experience that suffices for personal identity.) But his thesis also conflicts with the content-conservation version of the agency theory. For provided that someone has the right kind of psychological control she can survive over time even if there is significant diachronic variation in her practical objectives. See Stephen White, *The Unity of the Self* (Cambridge, Mass.: MIT Press, 1991), pp. 147-48. Rovane favors the addition of a more robust psychological condition that holds that a full-fledged agent must also be committed to “achieving overall rational unity within the set” of her projects (*The Bounds of Agency*, p. 164). I would still describe this as a causal condition, since it implies that something is an agent if its projects have the right sort of unity over time. But arguably this is a case in which the two versions of the agency theory begin to blur. It will depend on whether rational unity in projects is compatible with large variation in those projects over time.

\(^{16}\)David Shoemaker has recently argued that the “Parfitian picture can and does take into account the relation of authorial connectedness.” Shoemaker consequently rejects the suggestion that “Korsgaardian agents are importantly different from Parfitian persons.” See David W. Shoemaker, “Theoretical Persons and Practical Agents,” *Philosophy and Public Affairs* 25 (1996): 318-32, p. 327. If Shoemaker is right, this may explain the difficulty I have in classifying Brink’s view. In the passage quoted in the text, Brink makes the case for altruism by emphasizing that “experiences, beliefs, desires, ideals, and actions of each depend in significant part upon those of the others.” This suggests that a history of the right kind of psychological interaction—rather than diachronic comparisons in the contents of attitudes—is the relation that matters for personal identity. But Brink also claims to be sympathetic to Parfit’s analysis of personal identity.
persons. For the more basic issue is whether personal identity can plausibly be characterized as a relation that excludes physical and/or phenomenological continuity. What I shall argue is this. The continuity of agency is neither necessary, nor is it sufficient for personal identity. The argument for altruism from the unity of persons consequently fails because even if persons are not substances, there is a significant difference in the relation we bear towards the future occupant of our current body and the occupants of bodies by which we are separated in space.

3. Is the Continuity of Agency Necessary for Personal Identity?

People often say things like, “If I did not continue to identify with a certain social or political cause, then I would no longer be me, for that cause is part of my identity.” But while people often do speak this way, they do not usually have the concept of numerical or personal identity in mind on these occasions. Their use of the phrase “it would not be me” is a metaphor for a qualitative change in their personality. Similarly, the author who claims that she will live on through her books after her death is not using a single concept in a confused way. More likely, she is employing two different concepts of survival. These cases point to the more general observation: theories of personal identity address a very specific question pertaining to the conditions for the life and death of a person. The English language attaches diverse senses to the term “personal identity.” But since our aim is to arrive at a better understanding of the conditions for the numerical identity of persons, we should favor a method of inquiry that helps to separate these senses.

One such method involves inferring peoples’ beliefs about personal identity from their choices. The basic idea is that by eliciting preferences over carefully selected outcomes we can discover our convictions about personal identity. I call this the expressed preference method. The version of the method I favor owes its inspiration to Bernard Williams, and basically go as follows. You are presented with two future outcomes. Each contains a physically embodied subject who is related to you in a specific way. You can then choose which body is to be treated well and which body is to be treated poorly. Your preferences provide information regarding your views about personal identity. Strictly speaking, it does not matter what kinds of incentives are used in the examples. Whether the outcomes are aesthetically disagreeable or feature a large cash prize or physical torture is irrelevant. But physical torture is the incentive of

\[\text{17}^{\text{For related thoughts, compare Mark Johnston, “Human Concerns Without Superlative Selves,” in Dancy (ed.), Reading Parfit, pp. 149-79, at p. 174.}}\]
choice because nearly everyone can be presumed to dread the painful ravaging of what they believe will be their future body. So rather than first solving for their values, we can solve directly for their beliefs about personal identity.\(^{18}\)

Is the continuity of agency needed for personal identity? The notion that something must possess diachronic agency in order to be the same person is very popular among philosophers who work on the topic of freedom of the will. Arguably, there is some looseness in the way the term "person" is used in the context of the debate over freedom of the will.\(^{19}\) But some philosophers, who squarely address the problem of per-

---

\(^{18}\) Williams uses the expressed preference method to argue for a physical theory of personal identity. I believe that his conclusion is merely an artifact of the choice scenario that he constructs. By depriving people of their memories at the outset, he makes it impossible to determine whether a phenomenological unity could persist through the psychological changes he describes. Consequently, his argument only undermines the thesis that the continuity of agency is needed for survival. It does not test the adequacy of the neo-Lockean claim that phenomenological continuity is necessary or sufficient for personal identity. See his "The Self and The Future," reprinted in John Perry (ed.), Personal Identity (Berkeley: University of California Press, 1975), pp. 179-98.

\(^{19}\) In this vein Susan Wolf contends that: "We would have [if incompatibilists are correct] to stop thinking in terms that would allow the possibility that some lives and projects are better than others. Were we to make ourselves into the kind of creatures that ceased to think in these terms, we would lose the distinction between desires and values and, therefore, our distinction as valuing creatures. We would lose our ideals, our sense of self, and, I think, our status as persons. A world without reactive attitudes... would be a bleak, blank world of human brutes" (pp. 113-14, my emphasis). See her "The Importance of Free Will," Mind 90 (1981): 386-405, reprinted in J.M. Fischer and M. Ravizza (eds.), Perspectives on Moral Responsibility (Ithaca: Cornell University Press, 1993), pp. 101-18. Galen Strawson is eager to rebut Wolf's thesis that a life without reactive attitudes would be intolerably bleak. But on the question of whether non-agents can be persons he seems to agree with Wolf that they cannot. His thrust is that giving up on the first person singular is liberating rather than something to be feared. Strawson writes: "Suppose one did think about one's thinking in particular as a completely determined phenomenon. Then whatever thought one had, one would, pursuing the thought-experiment, think of that thought too as determined. And, this being so, no thought would ever be able to emerge as the true product of the familiar 'I', the putative true originator of thoughts, decisions, and actions that is not merely determined phenomenon: this 'I' would perpetually evanescs, however far one pursued the possible regress of thoughts about thoughts about thoughts." (80); "If there is, in a sense, no 'I', then there is nothing to fear in death and dissolution and there is no one there to feel fear in any case. For it is precisely the 'I's dissolution that is feared, and it is precisely the 'I' that does the fearing." See Freedom and Belief (Oxford: Oxford University Press, 1986), p. 98, n. 48, my emphasis (selections reprinted in Fischer and Ravizza (eds.), Perspectives on Moral Responsibility, pp. 67-100). All Galen Strawson references relate to the reprinted version. In a similar vein, Harry Frankfurt claims that "The essential nature of a person is constituted by his necessary personal characteristics ... They are especially characteristics of someone's will." See Necessity, Volition, and Love (Cambridge: Cambridge University Press, 1999), p. 113.
sonal identity, defend the view that continuity of agency is needed for being the same person over time.  

We can test that thesis in the following way. Suppose you are told the following. You shall soon be indoctrinated into a cult against your will using established brainwashing techniques. Assume that the process is so thorough that it irreversibly undermines the continuity in your plans, priorities, and other practical attitudes. After you have been indoctrinated, a body must then be tortured:

Option 1: The body you occupy now will later be tortured, and three innocent people will be spared torture.

Option 2: Some other body chosen at random will be tortured.

Assume that your choices are limited to Options 1 and 2. If you select Option 2, it will result in three extra persons being tortured at the conclusion of the sequence. (It can be supposed that three innocents will be plucked from the dungeons of Myanmar.) If you do not expect to survive the process of brainwashing, then the news that you’ve been given the choice described above should cheer you up a bit. The bad news, of course, is that the brainwashing will kill you. But that news is now mitigated by the fact that you have been given the opportunity to save three innocent people from a very disagreeable fate. If you lean even mildly in the direction of consequentialism, you should choose Option 1. However, if you believe that you can survive a thorough brainwashing, then Option 1 should terrify you. Someone who is a thoroughgoing rather than a mild consequentialist may still of course favor Option 1. A better test to elucidate the convictions of the thoroughgoing consequentialist would then eliminate the detail that three innocents are saved in Option 1. (So as not to stack the deck against Option 1, assume that your family will never learn about the tortures inflicted on your current body, or experience any psychic distress as a result of your choice.)

Let us assume that the brainwashing process is permanent. What will this be like? It seems reasonable to imagine the following. In the future

---

20For example, Korsgaard writes: “The price of complete practical normative scepticism, then, is nothing short of the loss of personal identity ... The final answer, then, to the question—what gives the instrumental principle its normativity?—is this: conformity to the instrumental principle is an essential part of what makes you a person.” See Christine Korsgaard, “The Normativity of Instrumental Reason,” in G. Cullity and B. Gaut (eds.), Ethics and Practical Reason (Oxford: Clarendon Press 1997), pp. 215-54, at p. 254, my emphasis. Compare Rovane: “persons have an internal reason to be concerned for their future well-being, at least insofar as such well-being rests on the fulfillment of their future desires. For ... anything that satisfies the condition for personal identity over time will have commitments to unifying projects that involve long-term activities ...” (The Bounds of Agency, p. 243, my emphasis).
"you" will be struck by the fact that people who "you" thought were your friends can't get used to "your" conversion to The Cause. (So as not to beg any questions, "you" refers in this context to the future occupant of the body you now occupy.) They will gradually desert "you." "You" shall have many new friends and projects in "your" life to take their place. You think these new friends and interests a waste of time now. But soon "you" will find them all very interesting. Apart from these changes, things will be pretty much the same: the future occupant of the body you now occupy will recall your past and all the salient episodes in your personal transformation. (If the future occupant does not recall these episodes, then phenomenological continuity is compromised and the example does not squarely test the proposition that continuity of agency is necessary for personal identity.)

If it is difficult to imagine oneself in this situation, then consider your husband or wife. After he has been brainwashed you continue to see him often. You recall all of your happy times together. But these occasions are tainted by your companion's obsessive belaboring of the point that you were both not really happy back then, and that he is far happier without you. It is certainly a terrible thing to contemplate one's partner in such circumstances. But it would be unspeakably awful to learn that your former companion is to have his body racked by excruciating pain. It would be no less dreadful to learn that one's body will be tortured. An enthusiasm for consequentialism notwithstanding, Option 1 is terrifying.

I have focused on the case of brainwashing because thorough and irreversible forms of brainwashing undermine continuity of agency, no matter which version of the agency relation we favor. When people are brainwashed the contents in their practical objectives normally undergo a dramatic change over time. But in addition, the psychological cause of that transformation conflicts with the minimal conditions for authorship on which Korsgaard and other agency theorists lay emphasis. So if the continuity of agency is genuinely necessary for personal identity, then Option 1 ought to be happily embraced by even mild consequentialists—once they have gotten over their grief for the brainwashing to come. Yet this seems incredible.

It may be urged in reply, however, that the agency relation comes in degrees, so while a brainwashed individual is no longer a full-fledged agent over time, he is agent enough to count as the same person. So the understandable aversion towards Option 1 fails to show that the unity of agency is not really needed for survival.

In searching for a decisive rejoinder to this reply, we are forced to

---

consider those exotic individuals who have relinquished their powers of agency in their totality. If “non-agents” can be the same person over time, then an aversion for Option 1 cannot be attributed to some lingering residuum of agency. The subjects in question should be physically and phenomenologically continuous so that the importance of the agency relation for personal identity is being tested. Could there be such an entity? Fortunately, the question has received some attention in debates over freedom of the will.

One argument against incompatibilism, pioneered by P.F. Strawson, goes as follows. Determinism is only threatening to our folkways if the reactive attitudes—judgments of guilt, shame, desert, admiration, and other attitudes that presuppose agency—can be extinguished. Our commitment to the reactive attitudes is, however, too deep to eradicate. So even if determinism is true this should have no effect on our determination to treat others and regard ourselves as agents. Yet some incompatibilists counter that it makes perfect sense for individuals to respond to the truth of determinism by suspending the entire range of their practical attitudes, adding that this suspension of agency is psychologically feasible. Let us suppose—as some incompatibilists insist—that individuals have the psychological resources to stop treating both themselves and others as agents. The key question this raises is this: Will ceasing to be an agent extinguish our future self-concern? Must it terminate the sense that some future outcomes are especially important because they implicate our self?

Presumably, the facts are these. As I come to see that all my initiatives are determined, I recognize that my opportunities to be creative and original have shrunk to none. Even worse, I realize that on those past occasions on which I felt a sense of achievement, these attitudes were also premised on a mistaken belief in my creative efficacy. Having reached this insight, I now suspend judgment about my activities. I no longer plan for the future. Nor do I deliberate in the expectation that I can thereby stand above the forces that determine my actions. Nor do I reflect on the quality of my desires. For each act of practical deliberation is itself caused. Ultimately, I come to accept that there is no creative force behind my actions. I lose all desire to exercise my powers of agency because I no longer see any point in that exercise.

This insight into my condition may be disappointing. But it is surely

---


23This is argued by Galen Strawson, who adds that Buddhist meditation can facilitate the dissolution of agency (Fischer and Ravizza, p. 68).
not the end of my world. For I continue to recall the details of my life as it has unfolded up until now. I may no longer aim to remember the past, but memories, like fantasies, do not always depend on acts of will. My other cognitive capacities remain intact. In addition, I retain impulses of a more practical sort. When I am hungry I eat, when thirsty I drink. For these desires are not sustained by the thought that they were freely chosen. Like my aversion to being scalded or frost-bitten, there was never any illusion that I was a creative and self-determining force in selecting these goals. So there is no stool to be kicked out from under these passionate states when I acknowledge the truth of determinism.

We can call these unchosen wants and aversions natural desires. It is true that these natural desires can often be satisfied in different ways. And since I am no longer an agent, I have no interest in deliberating on the merits of these ways. But when different modes of satisfaction are available—such as a choice between eating with my fingers, chopsticks, or utensils—I simply defer to social custom. For by deferring to custom I can avoid having to weigh the merits of rival practices while still attending to my wants.

In supplying a phenomenology for an individual who has relinquished her agency, I have relied on the Pyrrhonian skeptics. The Pyrrhonians professed to find this lifestyle very attractive. By contrast, when I imagine what it is like to be a non-agent, I believe that my life will have become much less interesting, in much the same way that it would be less interesting if I were paralyzed, color blind, or deaf. It nonetheless seems bizarre to conclude that persons cannot survive this transformation. For they retain the cognitive abilities—including powers for imagination and fantasy, and other signs of intelligence—which distinguish people from quintessential non-persons. They furthermore retain their natural desires. Turning to the first person perspective, when I imagine life as a non-agent, I remain strongly attracted to Option 2 (unless I am in a mood for self-sacrifice). For I retain an unshakable, natural aversion to the painful destruction of my body. This repulsion to Option 1—and the personal concern for the future it signifies—cannot be reconciled with the proposal that continuity of agency is necessary for personal identity.

The reply that a bona fide non-agent could not form preferences over the two outcomes I described may miss the point. The imposition of a choice framework is a useful device for bringing our intuitions about personal identity into focus. But it is only a heuristic device. The choice between options can be eliminated. We can simply investigate whether the torture in the future of the body we now occupy is a source of anxiety. If the prospect of first undergoing brainwashing, or otherwise giving up our powers of agency, makes us no less concerned for the future of that body, then we should conclude that the continuity of agency is not
needed for personal identity.

The objection might continue that any entity that has preferences for outcomes taken in isolation must be an agent. There are two things to be said in response. First, this complaint raises definitional issues about the nature of agency. Yet if simply having preferences over outcomes is enough to qualify something as an agent—and by extension, a person—then it must at least be true that thicker forms of agency, such as Korsgaard’s authorial relation, are not needed for personal identity.

Second, the question of how a Pyrrhonian—or some other non-agent—might react to the thought experiment described in the text is complex. My sense, however, is that there is no deep incompatibility between forming preferences over outcomes and the relinquishing of agency. The Pyrrhonians, for instance, took pains to emphasize that the suspension of judgment involved merely giving up evaluations. This suspension of judgment then promotes tranquility because when misfortune strikes, skeptics do not form the judgment, “something bad has happened to me.” So their misery is not compounded by their beliefs about value. But skeptics do not deny the reality of the impulses and aversions we are endowed with by nature. These natural impulses, along with deference to custom, furnish a basis for choice.

An exhaustive investigation of the necessity claim would not stop here. We would need, for example, to examine our considered convictions from a third-person perspective. If we encountered a community of Pyrrhonian skeptics, or some other highly impractical human beings, would we resist viewing them as persons—and perhaps treat them like mere animals? The necessity claim merits a more thorough investigation than I shall give it. But for our current purposes, the thesis that must be gotten to grips with is the view that the continuity of agency suffices for personal identity. For it is only if the sufficiency claim is borne out—singly or in conjunction with the necessity claim—that the argument for altruism from the unity of persons can succeed.

4. Does Continuity of Agency Suffice for Survival?

Let’s refresh our memories regarding the parameters of the discussion. The agency relation can be characterized in a variety of ways. But if the

---

argument for altruism from the unity of persons is to succeed, then the
continuity of agency must exclude certain elements. It can't require
physical or phenomenological continuity, because we are simply not re-
lated to other people in those ways. Consequently, in testing the suffi-
ciency claim we must probe our attitudes towards a very specific kind of
future. It should contain persons whose plans, deliberative principles, and
other psychological states connected to agency are suitably related to our
own. But the future should not contain individuals who possess our cur-
rent brain and/or body, or individuals who directly recall our experiences.

We can once again investigate our convictions by eliciting our reac-
tions to a carefully constructed example. The example contains fictional
elements. But it does not incorporate features that are incompatible with
our existing technologies. We could be compelled to make choices of
this kind within our lifetimes. It therefore seems idle to complain that we
have no solid intuitions about cases of this kind (a complaint sometimes
made about science fiction examples in the personal identity literature).

You receive the following promotional material in the mail:

Dear Sir/Madam,

Congratulations! You have been pre-approved to become a value do-
nor. Gerital is a world leader in the business of child education, operat-
ing primary schools in over fifty developing countries. Our team of ex-
erts will conduct comprehensive interviews gathering information about
your life, works, and your personal ethical outlook. Sample: "Should you
lie to friends? Here are the most important reasons why not." (Don't
worry if you are unsure about what you believe. We will provide a team
of expert consultants to clarify your ethical thoughts.) This information
will then form the basis for a comprehensive school curriculum in one of
our client states. We guarantee exposure of your personal ethical outlook
to over 100,000 young minds for the duration of their all-important for-
native years!!! To qualify as a value donor you must simply agree to
transfer your assets, including your brain and body—to be killed pain-
lessly according to the method of your choice—to Gerital. Isn't it time
you got serious about longevity?

The Favor of a Reply is Requested.

If the transmission of values, priorities, and other agency-related
properties is the unity relation for persons, then this offer should be at-
tractive. Yet it is hard to take seriously the notion that the inculcation of
our priorities in others—even enormous numbers of others—is adequate
compensation for the loss of physical and/or phenomenological contin-
uity. Or so I believe. And one can probably do no better than explain the
sanity of one's belief in the context of the debate over the nature of per-
sonal identity. Let me therefore consider some potential misgivings.

One worry about the example may be that it stacks the deck against the agency theory because the grounds for rejecting the company's offer have nothing to do with survivalist concerns. Rejection of the offer is prompted by ethical considerations. In this vein, some may fear that the agreement is unattractive because it threatens to compromise the quality of many children's education. Other reservations may be grounded in concerns over one's beneficiaries. What will happen to them if your estate is ceded to Gerital? Still other worries may be aesthetic: What will happen to your body after it is dead?

Yet these fears don't scuttle the example. At most they show that details must be added. We can assume that the schools administered by Gerital are about as good as the available alternatives. The governments of these client states are democracies accountable to their people, who have endorsed this commercial transaction in national referenda. You have no dependents. The body will be cremated with its internal organs intact, and so forth. In fact, we should add whatever details we like to dispose of the worry that our decision is being guided by practical considerations.

My sense is that adding these details has no tendency whatsoever to render a trade of our body for value donor status any more palatable. I suspect that any conflicting sentiments are grounded in considerations that have nothing to do with personal identity. More specifically, becoming a value donor may be attractive because it is a way to consummate our plans. And the prospects of this success may on occasion outweigh a preoccupation with survival. If you are deeply committed to preserving the environment (say) then having a lot more preservationists on hand is a way to succeed, and success in that plan may matter more to you than your life. But success should not be confused with survival. Some people can have their plans consummated posthumously. While others can fail in their plans and still live on—having to endure the barbed description "lifelong failure." Being successful is one thing, survival or persistence over time is another. We should therefore neutralize the "success variable" in our example. We should stipulate that becoming a value donor has no tendency to raise or lower the probability that you will succeed in realizing your plans. When the success variable is neutralized, then one lingering reason for becoming a value donor gets extinguished.

One further response to my argument holds that, far from undermining the agency theory, our reluctance to become value donors under these circumstances supports the opposite conclusion, namely, that continuity of agency suffices for personal identity. It may be alleged that the reason we reject becoming a value donor under these circumstances is that our
bodies currently host a perfectly viable agent, and this agent would be extinguished by the terms of the agreement.

Accordingly, we have two hypotheses to account for our refusal to take up the offer to become a value donor. This reluctance can be explained by the conviction that physical and/or phenomenological continuity is what matters for survival. Or it can be accounted for by the thesis that the continuity of agency suffices for survival in conjunction with a further principle that I'll call the concentration principle. This principle holds that concentrated forms of agency—housed in a single body for example—matter more for survival than the diffusion of our plans and values to other minds. Concentrated agency is more valuable for survival than continuity of agency extending to other minds—even very large numbers of other minds.

To test these rival hypotheses, suppose that letter from Gerital is preaced with the following statement:

*You have been selected to participate in our value donor program because a recent hospital test has revealed that your body will die from health complications within two years.*

If the continuity of agency genuinely suffices for survival, then the offer to become a value donor should become ever more attractive as the lifespan for your current body contracts. For as your body’s anticipated lifespan shrinks, so does the viability of your agency in its concentrated form.

Does information of this kind have any impact on our decisions? I can report that in my own case the information that my current body will perish does nothing to entice me to part with it any sooner. I expect that others will feel the same—at least if they have made the effort to vividly imagine the details of the situation.

Suppose this correctly describes our reactions. A proponent of the agency theory might fall back on the response that intervals of concentrated agency—even very limited intervals—are more important for survival than more diffuse agency relations. And indeed, someone who subscribes to an agency theory of personal identity could resort to this desperate defense. But it should be clear that the entire strategy of privileging concentrated forms of agency at the expense of diffuse varieties remains at cross-purposes with the argument for altruism from the unity of persons. For it implicitly concedes that the relation we bear towards the future occupant of our current body normally differs in a significant way from our relation to minds located in other bodies. Reflections on the nature of personal identity would therefore confirm rather than undermine a preoccupation with the body and its limits.
5. Conclusion

Let me conclude with a diagnosis. The attractiveness of the agency theory of personal identity stems from the choice of examples that invite confusion between a concern for survival and other objectives.

When we are given the information that our current deliberations will extend into the future, it is natural for us to assume that there will be a unified stream of consciousness over time, a stream to which we have ongoing access. For planning, choice, and other mental activities related to the continuity of agency all have phenomenal properties. This is conveyed in a host of metaphors. Expressions such as "agonizing over a decision," "being in the grips of an obsession," "deliberating in a cool hour," "being hot tempered" are familiar phrases that attest to the fact that practical deliberation has experiential properties as palpable as physical sensations. Unsurprisingly, someone who anticipates continuity in her deliberative processes also anticipates phenomenological continuity.

To test the proposition that the continuity of agency suffices for identity, we must then take pains to alter the default assumption that phenomenological continuity accompanies continuity of agency. The easiest way to do this is by considering cases in which the agency relation extends across bodies separated by space. When those cases are considered, the suggestion that continuity of agency suffices for personal identity becomes unpalatable.

A similar point is raised by appeals to our habitual concern for our children, clubs, and other groups that share our practical priorities. These patterns of concern are frequently taken as evidence in support of the conclusion that continuity of agency is connected to survival. We care about our children because we survive in them, the story goes. But this psychological evidence is irrelevant, since those same patterns of concern can obviously be accounted for in a variety of ways. We may care that our children have our values because we believe that these values make for a happy life, and we want our children to be happy. We are concerned for the associations to which we belong because we are friends with their members, or value the practical objectives that those associations promote. These, and other explanations, compete with the very selective hypothesis that our interpersonal concerns stem from a preoccupation with survival.

Throughout this essay, I have appealed to our reactions towards various thought experiments. These examples are certainly not the only—and are arguably not the best—specimens for testing the agency theory of personal identity. But it seems clear that if our reactions are to play any part in the debate over personal identity, they will have to be reactions towards carefully controlled examples of this kind. Independent variables
must be isolated and the remainder neutralized.

The conclusion that emerges from this controlled mode of inquiry is that the conventional boundaries attributed to the self are not simply a hangover from some defunct religious or metaphysical theory. Those boundaries flow from material factors, chiefly, the observation that our minds normally only have access to the experiences hosted in our brains. Persons, of course, are more than mere experiences. A mature human being typically combines those experiences through memory and imagination into a richly textured mental life. It is the power to weave these rich textures that matters to us. It is in virtue of these cognitive powers that persons are distinct from mere human beings and other animals. It is our continued access to this stream of experiences, and the anticipation of more to come, that grounds our expectation of surviving to the future. But it is the brains in our heads that furnish our normal mode of access to the experiences we have at first hand. This fact of human biology explains why efforts at rendering morality attractive through a denial of the separateness of persons will always seem fishy. The boundaries of the body are salient because they normally correspond to the psychological relations—in the form of the continuity of consciousness—that matter for personal identity.

My preferred alternative is to accept the view that the separation between persons is deep, while accommodating it within our conception of practical reason. With regard to the "Why be moral?" or the "Why be altruistic?" question, the mode of accommodation I favor holds that moral reasons are autonomous. Their warrant as reasons has nothing to do with prudential grounds. Whether this argument can be successfully pursued remains to be seen. It may be that at the end of the day we shall have to settle for something less than the traditional view that all persons have reason to care about others. I hope that external reasons for altruism can be vindicated. But I doubt that an attempt to redraw the boundaries of the self will contribute to altruism's prospects.25

Sam Black
Department of Philosophy
Simon Fraser University
Samuel_Black@sfu.ca

---

25I would like to thank Carol Rovane and an anonymous reviewer from this journal for very helpful comments that shaped the evolution of this paper.