

Personal Identity and Practical Reason: The Failure of Kantian Replies to Parfit

JONNY ANOMALY *University of Virginia*

*ABSTRACT: This essay examines and criticizes a set of Kantian objections to Parfit's attempt in *Reasons and Persons* to connect his theory of personal identity to practical rationality and moral philosophy. Several of Parfit's critics have tried to sever the link he forges between his metaphysical and practical conclusions by invoking the Kantian thought that even if we accept his metaphysical theory of personal identity, we still have good practical grounds for rejecting that theory when deliberating about what to do. The argument between Parfit and his opponents illuminates broader questions about the relationship between our metaphysical beliefs and our practical reasons.*

*RÉSUMÉ : Cet article examine et critique un ensemble d'objections kantiennees à la tentative de Parfit, dans *Reasons and Persons*, d'ajuster sa théorie de l'identité personnelle à la rationalité pratique et à la philosophie morale. Plusieurs des critiques de Parfit ont essayé de rompre le lien qu'il tisse entre ses conclusions métaphysiques et pratiques en évoquant l'idée kantienne selon laquelle, même si nous acceptons sa théorie métaphysique de l'identité personnelle, il existe cependant de bonnes raisons pratiques de rejeter cette théorie lorsque nous délibérons à propos de ce que nous devons faire. Le débat entre Parfit et ses adversaires nous éclaire sur un questionnement plus large à propos du rapport entre croyance métaphysique et raison pratique.*

Dialogue XLVII (2008), 331-50

© 2008 Canadian Philosophical Association / Association canadienne de philosophie

In *Reasons and Persons* Derek Parfit raises important questions about the relationship between our metaphysical beliefs and our practical reasons by exploring the consequences of his reductionist theory of personal identity for ethics and rational choice. Near the end of *Reasons and Persons* it becomes clear that one of Parfit's central concerns is to challenge John Rawls's influential claim in *A Theory of Justice* that utilitarianism is an implausible moral theory because it makes a metaphysical mistake: in its concern to maximize the aggregate utility of persons, it fallaciously applies individual maximizing rationality to groups of distinct persons (Rawls 1999, pp. 25-26). In effect, Rawls suggests that the fatal flaw of utilitarianism is not its counter-intuitive ethical consequences, but rather the fact that it ignores a deep metaphysical distinction between persons. In contrast, Parfit thinks that once we accept his reductionist theory of personal identity, the boundaries we erect between ourselves and others will begin to crumble, thus increasing our willingness to endorse impersonal, utilitarian moral principles: "the Utilitarian View may be supported by, not the conflation of persons, but their partial disintegration" (Parfit 1984, p. 336).¹

Rather than address Parfit's views on utilitarianism, I shall defend his more general assumption that theoretical insights about personal identity may legitimately alter either the substance or scope of our practical reasons. My argument takes a conditional form: if we accept Parfit's theory of personal identity, then we are committed to certain practical consequences. In order to substantiate the antecedent, I begin with a brief overview of Parfit's reductionist view of personal identity and his consequent view that what matters in survival is not numerical identity but rather the persistence of certain physical and psychological relations. I then begin the main task of the article. The goal is to defend Parfit from a pair of critics—both of whom acknowledge some debt to Kant—who think we can accept Parfit's metaphysics while blocking the practical inferences he wishes to draw. Both critics exploit the Kantian belief that we may view a particular event from either a practical standpoint or a theoretical standpoint, and arrive at two different yet equally legitimate conclusions.² Although I refer to such views as "Kantian," I make no attempt to interpret Kant's writings on personal identity.

1. Debunking the "Standard View"

According to what Parfit calls the "standard view," persons are physical objects of a certain sort, and personal identity just consists in a particular person tracing a unique spatio-temporal path. On Parfit's reductionist view, by contrast, since persons are best described in psychological, rather than physical terms, the main ingredient of personal identity is the persistence of certain psychological relations over time.³ Parfit thinks physical relations are important to personal identity only to the extent that a phys-

ical medium is necessary to support the kinds of mental events that comprise persons.

Parfit devises a series of thought experiments to cast doubt on the standard view. In one of his more familiar thought experiments, *The Branch-Line Case*, we are asked to imagine the advent of Star Trek-style transportation. A “Teletransporter” would scan and record your mental and physical state at a particular time—including the positions and interrelations of all your brain and body cells—and then use this information to transport you somewhere else by re-creating your body and brain out of new material. The Teletransporter would then either destroy your original body or else preserve it. If the original were destroyed, Parfit says, many people after careful consideration would be inclined to call the replica *different* from the original body but *identical* to the original person: although it is composed of fresh atoms, the replica is qualitatively identical to and psychologically continuous with the original person. Someone contemplating whether to Teletransport might thus anticipate travel rather than death. If, however, the standard view of personal identity were correct, we should anticipate death rather than travel, since the spatio-temporal path of the person in question is interrupted. Thus, if the standard view of persons were correct, the advent of Teletransportation would amount to a new way of committing suicide, rather than a new way of travelling.

The assumptions underlying the very idea of Teletransportation are not trivial. It is first assumed that persons are connected with but conceptually distinct from their bodies (though not in the way Descartes believed).⁴ But it is also assumed that the medium in which a particular person is instantiated can be altered or even replicated while preserving the psychological relations Parfit considers crucial to survival. Both assumptions may be doubted (though I shall not defend them here, since my main objective is to explore the practical consequences of Parfit’s view). Moreover, even if each assumption is granted, the pair of assumptions may seem inconsistent: if we manipulate or duplicate the medium, we apparently alter the person who goes with it. But this does not follow. If a person fundamentally *is* certain kinds of psychological relations, and if we can conceive of psychological relations as dependent upon but distinct from the physical medium which supports them, then we can coherently imagine the possibility of scanning and copying a functioning brain so that its memories, beliefs, etc. are recorded and then re-instantiated in a new medium.

2. Continuity and Connectedness

If we consider a case of Teletransportation in which both the replica and the original person survive, we immediately face a conceptual problem for personal identity. Although both survivors are exactly like the person

scanned, only one of them can be *numerically* identical to the original person. This kind of case suggests that what fundamentally matters in survival is not numerical identity, but rather the persistence of certain psychological traits that we care about. Parfit distinguishes two kinds of psychological relation that he thinks fundamentally matter in questions of survival: connectedness and continuity.

Psychological Connectedness is the holding of particular direct psychological connections.

Psychological Continuity is the holding of overlapping chains of strong connectedness.

Examples of a *psychological connection* between person P at one time and person P' at a later time are an *intention* that is formed by P and then carried out by P', an *experience* had by P and then remembered by P', and certain distinctive desires and beliefs integral to the psychological make-up of P which persist in P'. It is important to understand that for Parfit a particular psychological connection—such as the intention I recently formed to write this article, and the current fulfillment of that intention—either holds or does not hold.⁵ Thus, although Parfit confusingly says that psychological connectedness “admits of degrees” and can be “strong” or “weak” (1984, pp. 206, 304), he does not mean that some particular experience memories, for example, are more vivid than others, but rather that the number of connections between P and P' may be greater or lesser. To use a rope rather than a chain metaphor, the “strength” of a Parfitian rope of connectedness does not refer to the quality of thread it contains, but instead to the number of intertwining strings which comprise it.

Psychological continuity consists in “overlapping chains” of strong connectedness in the way in which the strands in a rope mesh together (*ibid.*, p. 222). If many distinct psychological connections hold between person P at t_1 and P' at t_2 , then P' is *continuous with* P. If we imagine a single strand which ties a particular memory had by P' to an earlier action of P, then P and P' exhibit some *connectedness* but not necessarily any *continuity*. If many strings of experiences and subsequent memories of those experiences bind P to P', or many strings of beliefs and desires persist between P and P', then P and P' are *psychologically continuous* with one another.

Psychological connectedness and continuity jointly constitute what Parfit calls “Relation R.”⁶ According to Parfit, Relation R is what matters in survival. The only difference between personal identity (“PI”) and a full realization of Relation R (“R”) is that personal identity (*qua* numeric identity) implies Uniqueness (“U”). More schematically, on Parfit’s reductionist view of personal identity, $PI = R + U$. In most cases, U has no

intrinsic importance (unless you have a special attachment to the particular atoms that currently make up your body and brain). Thus, on Parfit's reductionist view, what matters in survival is that the maximum degree of R obtains between P and P'.

3. Persons and Reasons

If, as Parfit says, the standard view of personal identity is both widely believed and false, this would apparently have far-reaching consequences for our practical reasons and moral commitments. For example, if my initial reason for setting aside money for retirement depends on the assumption that I will be identical to the retired person with my name, and if I subsequently discover that I will eventually have little if any psychological continuity with that future person, my original reason seems to vanish. It might also seem that if we can expect diminishing psychological continuity over time, we would have less reason to draw a sharp distinction between our own future welfare and the welfare of others in the future.

According to some commentators on an earlier work (Parfit 1973), Parfit suggests that "considerations about the nature of personal identity may provide some support for utilitarianism" (Scheffler 1982, p. 231). This is true but misleading. Parfit does think a reductionist view of PI may help defeat certain objections to utilitarianism and that it may help explain why some people accept utilitarianism. He does not, however, claim that reductionism suffices to justify any particular moral view. Other commentators suggest that since Parfit thinks of personal identity in terms of Relation R, he is committed to the claim that "we should identify with and care about ourselves and each other [only] to the degree to which our parts are connected by Relation R" (Wolf 1986, p. 710). In fact, however, Parfit thinks reductionism has more modest practical consequences. First, he claims that "it can be rational to care less [about my own future] when one of the grounds for caring will hold to a lesser degree" (1984, p. 313). If my special concern for my own future is grounded in Relation R, then this rationally licenses, though does not require, me to act on the basis of "R-variant" concern. As Parfit puts it, "my concern for my future may correspond to the degree of [R] between me now and myself in the future" (*ibid.*).⁷

Parfit also suggests that we may take similar attitudes toward other people: we may justifiably adjust our attitudes toward others to account for diminishing degrees of Relation R. For example, my previous special concern for my primary schoolmates may reasonably decline as we "grow apart" or become less R-related to our earlier selves. For similar reasons, Parfit thinks we may rationally change our attitudes toward malefactors and promise-breakers: "Reduced psychological connectedness would reduce both responsibility for past crimes, and obligations to fulfill past commitments" (*ibid.*, p. 446).

4. Korsgaard on Metaphysical Theories and Practical Reasons

Christine Korsgaard denies that Parfit's reductionist view of personal identity has any practical consequences. Her grounds are broadly Kantian, and her strategy is to provisionally accept Parfit's view of PI as a metaphysical theory but to deny its alleged implications for rational choice. She begins by conceding that "Parfit has established that there is no deep sense in which I am identical to the subject of experiences who will occupy my body in the future" (1989, p. 109). She then sets out to defend her view: "I nevertheless have reasons for regarding myself as the same rational agent as the one who will occupy my body in the future," adding that "these reasons are not *metaphysical*, but *practical*" (*ibid.*, p. 109). Ultimately, Korsgaard hopes to cleave the connection Parfit makes between personal identity and practical reasons by undermining the authority of any purely metaphysical account of PI.

Before turning to her central arguments against Parfit's reductionist view of PI, I should highlight an important mistake Korsgaard makes in interpreting Parfit's view about the practical consequences of reductionism. She says, "Parfit assumes that my attitude about the desires of the future inhabitant of my body should be based on the metaphysics of personal identity. . . . But this argument from the metaphysical facts to normative reasons involves a move from 'is' to 'ought' which requires justification" (1989, p. 112). In fact, Parfit makes no such assumption and no such move.

In discussing the practical consequences of his reductionist view of PI, Parfit distinguishes two possible implications, neither of which involves a move from "is" to "ought." According to the first, which he calls the Extreme Claim, reductionism shows that we have no reason at all to be concerned about our own futures (1984, p. 307). Only if personal identity were based in some further fact beyond Relation R could we have any such reason (*ibid.*, p. 311). The Moderate Claim, by contrast, which Parfit favours, is that the reductionist view *can* justify our taking a special interest in our own futures, given a sufficient degree of Relation R (*ibid.*, p. 313). The Moderate Claim does not imply that we are rationally required to proportion concern for our future selves to the degree to which Relation R is expected to obtain. It merely implies that Relation R *may* play a legitimate role in deciding how much interest to take in a future self. Nevertheless, this appears to confirm Korsgaard's accusation: Parfit moves from a factual claim about personal identity to a normative claim about rational choice. But this is not quite right. Parfit's move is better seen as a metaphysical adjustment to the *scope* of foregone practical reasons.

Assuming with Parfit that we each have reason to act in ways that promote our own good, a proponent of the "standard" or non-reductionist

view of personal identity might claim that we ought to act in ways that are neutral between promoting our own good now and our own good in the future. The Extreme Claim simply suggests that since the standard view is false, we have no special reason to care about our future, while the Moderate Claim suggests that since the standard view is false, we might ground concern for future selves in Relation R rather than personal identity. According to Parfit, both claims are rationally defensible (*ibid.*, p. 312). More importantly, neither claim violates Hume's Law—neither moves from *is* to *ought*. Since both claims take normative reasons as inputs and transform them into (different) normative outputs, both the Extreme Claim and the Moderate Claim move from *ought* to *ought*. They differ only in the direction in which they allow similar metaphysical facts to channel normative reasons.

5. Korsgaard and the Unity of Agency over Time

Korsgaard gives two arguments against Parfit's reductionist view and for the necessary unity of agency over time, or personal identity, both of which appeal to the idea that personal identity is forged through practical deliberation and action. Her first argument begins with the premise that most of what we care about takes time. This includes friendships, personal projects, and careers. Korsgaard maintains that "in choosing our careers, and pursuing our friendships and family lives, we both presuppose and construct a continuity of identity and of agency" (1989, p. 113). This suggests a transcendental argument, but one which concerns the preconditions for certain kinds of action rather than experience. It can be reconstructed as follows:

- (1) In order to engage in long-term relationships and pursue projects, we must have (construct) an identity that spans a significant tract of time.
- (2) We have such relationships and pursue such projects.
- ∴ ---
- (3) We must have (construct) an identity that spans a significant tract of time.⁸

The argument is deductively valid, but either unsound or impotent, depending on how we interpret its premises. Beginning with the second premise, it is noteworthy that personal identity theorists find marginal and counterfactual cases significant precisely because they do not conform to the normal pattern "we" exhibit. Premise (2), however, ignores these cases. Moreover, it ignores the familiar fact that projects and friend-

ships can be interrupted by personality changes or life circumstances beyond our control. Nonetheless, premise (2) is true as an empirical description of most people most of the time. Premise (1), meanwhile, seems to be a conceptual platitude about relationships and projects.

Let us, then, provisionally grant both premises in order to focus on the conclusion Korsgaard draws from the argument. Korsgaard takes the conclusion to show that personal identity spans a normal human life span—that “you are one continuing person because you have one life to lead” (1989, p. 113). Yet the argument she suggests above cannot generate this result. This is because premise (1) is plausible only if “significant tract of time” is taken to mean less than or equal to a normal biological life span. After all, a meaningful relationship can exist over several months or years, and a rewarding project can span a few weeks. Neither relationships nor projects necessarily last indefinitely long, nor do we always intend them to. Even when they do last much of one’s life, they may persist in the face of significant changes to those involved. The pursuit of projects and persistence of friendships requires some degree of continuity in the people involved, but it does not require that the content of the project or the nature of the persons involved remain identical over time.

Thus, if Korsgaard’s premises are interpreted in a plausible way, her argument does not show anything about personal identity in unusual cases or over long tracts of time. It only shows that having relationships and pursuing projects takes time. To this extent, it is compatible with Parfit’s view. If, however, it is supposed to show (as she suggests) that projects and friendships are the glue that bind together the temporal stages of an entire human life, then it is not only incompatible with Parfit’s view, but it rests on the false assumption that friendships and projects cannot exist unless they span an entire lifetime.

In her second argument for the necessary unity of agency over time, Korsgaard supposes for the sake of argument that a succession of rational agents occupies her body (*ibid.*, p. 113).⁹ In this case, she claims, each agent will need to cooperate with other temporally separated agents in order to advance their own ends—a sort of diachronic, intrapersonal version of Hobbesian contractualism. Given a shared desire to advance their own ends, each self in a string of selves ought to engage in mutually beneficial cooperation (much like Hobbesian agents ought to act in ways that bring about mutual benefit through social stability). Korsgaard quickly realizes that the analogy is strained. Although a current self might benefit a past self by carrying out its plans, it has no reason to do so if the plans are inimical to its current interests. Moreover, on the assumption that selves are discrete entities, there is no reason for a current self to ensure the well-being of a future self that does not yet exist, for, if her grounds for helping a future self were for mutual benefit, she would be committed

to an erroneous belief in backward causation—helping a future self in the expectation of some current benefit.

Perhaps a similar insight leads Korsgaard to refine her second argument. Even on the assumption that your body is occupied by successive selves, Korsgaard claims, “your present self must identify with something from which you will derive your reasons, but not necessarily with something present” (*ibid.*, p. 113). For any “deliberative choice” or plan you make concerning your future, that plan will very likely “carry” you some distance into the future. Your present self must identify some future self—which it regards as identical with its current self—as the future beneficiary or executor of a current plan (*ibid.*).

With this conceptual claim I have no quarrel. But it is not clear how it damages Parfit’s reductionist view of personal identity. Korsgaard is correct that practical deliberation or planning always concerns future events, and that forming and executing plans involves the presumption of identity (or at least a high degree of Relation R) between the person who formed the plan and the future person deciding whether to execute it. But she erroneously concludes from this claim that we must in fact be unified agents over indefinitely long tracts of time. This is wrong for two reasons. First, as Korsgaard would agree, merely forming a plan does not guarantee that it will be carried out. Our plans might change or the world might prevent us from carrying them out. Second, and more importantly, even if plans are carried out in the future, the reductionist view does not imply that cooperation and consequent unification of successive selves has taken place. This is because, even on the reductionist view, it does not follow from the fact that the plans of a self at t_1 are carried out at some point in the future t_2 that the self which carries out the plan is a different future self. Selves are not necessarily so transitory, except on the most extreme reductionist view of personal identity. Parfit accordingly warns that “talk about successive selves can easily be misunderstood, or taken too literally. [This way of talking] is suited only for cases where there is some sharp discontinuity, marking the boundary between two successive selves” (Korsgaard 1989, p. 306). So the modified version of Korsgaard’s second argument can only derail an extreme reductionist view that seeks to show that each ordinary person is merely a long series of temporally distinct selves with little if any relation to each other. And this view is indeed implausible. Let us nevertheless grant Korsgaard’s second argument against Parfit’s reductionist view and see what follows.

Korsgaard claims the upshot of her argument is that “there is a necessary connection between agency and unity which requires no metaphysical support” (*ibid.*, p. 115). To justify this anti-reductionist conclusion, Korsgaard invokes the Kantian thought that we may think about human agency and personal identity from either of two legitimate standpoints. According to Korsgaard, the theoretical standpoint suggests that we are

mere bundles of psychological phenomena which become less connected over time. But from the practical standpoint we rightly see ourselves as unified agents over time. Korsgaard explicates her use of the two-standpoints strategy with an analogy between the problem of personal identity and the problem of free will:

We may regard ourselves as objects of theoretical understanding, natural phenomena whose behavior may be causally explained and predicted like any other. Or we may regard ourselves as agents, as the thinkers of our thoughts and the originators of our actions. . . . As objects of theoretical study, we see ourselves as wholly determined by natural forces, the mere undergoers of our experiences. Yet as agents, we view ourselves as free and responsible, as the authors of our actions and the *leaders* of our lives. The incongruity need not become contradiction, so long as we keep in mind that the two views of ourselves spring from two different relations in which we stand to our actions. . . . These two relations are *equally legitimate*, inescapable, and governed by reason, but they are separate. (1989, p. 120; emphasis added)¹⁰

Korsgaard's analogy between how we might view our actions (*qua* free vs. determined) and how we might view the unity of our lives over time (*qua* reducible to psychological phenomena vs. unified) is entirely appropriate. From the inside, or the deliberative standpoint, we certainly seem to make free choices and live a single, unified life. But might this appearance be deceptive?

6. Korsgaard's Compatibilism

Korsgaard argues that taking a practical, deliberative view of ourselves is "equally legitimate" to taking a theoretical, explanatory view: the facts that each standpoint reveals are equally facts. This compatibilist view of personal identity is intuitively plausible, but it faces significant epistemological problems. The first problem is that it implies that a particular person-stage can be given contradictory descriptions, both of which are equally true. To illustrate the problem, let us define "p" as the proposition that Pablo Picasso at age seventy-five is identical with Pablo Picasso at age twenty-five. Korsgaard's compatibilism implies that from the practical standpoint "p" may be true despite the fact that from the theoretical standpoint " \sim p" is true. Thus, "p & \sim p" are compatible—Picasso at age seventy-five both *is* and *is not* identical to Picasso at age twenty-five.

In Korsgaard's defence, we might avoid inconsistency in simultaneously believing "p & \sim p" by appropriately relativizing truth claims. One way of doing so is to claim that when we deliberate practically, we discover that p is true, but when we deliberate theoretically, we find that \sim p is true. If so, "p & \sim p" may be simultaneously true. But this is a pyrrhic victory. If truth is merely perspectival—i.e., relative to one of two equally legitimate stand-

points which we may occupy—then we can never describe either ourselves or the world as they really are. Although this version of the “two-standpoints” argument is coherent, it dogmatically limits the pretension of our theories. It implies that our ideas about biology and chemistry, for instance, are deeply dependent on the standpoint we take. Worse still, perspectivism cannot hope for a ground-level justification, since its truth precludes us from making non-perspectival claims. An opponent of this version of the two-standpoints strategy need not assume that we can certainly describe the world as it really is. He need only reserve the metaphysical right to try—a right which perspectivism forbids us from exercising.

A more compelling way for Korsgaard to countenance inconsistent beliefs generated by the two-standpoints argument is to distinguish the justificatory standards of practical beliefs from those of theoretical beliefs. If it is theoretically rational to believe *p*, but practically rational to believe $\sim p$, then perhaps the belief set “*p* & $\sim p$ ” can be justified. To see how, consider the following distinction:

- (1) *p* is the rational thing to believe.
- (2) believing *p* is the rational thing to do.

Robert Nozick illustrates the distinction with the case of a doting mother presented with courtroom evidence that her son has committed a terrible crime.¹¹ Nozick suggests that although concluding that her son is guilty may be the rational thing for her to believe, denying his guilt or abstaining from forming an opinion on the subject (if this is psychologically possible) may be the rational thing for her to do. This is so if her practical goal of avoiding deeply painful beliefs outweighs her theoretical goal of increasing her share of justified beliefs.¹²

According to the present proposal, Korsgaard faces a similar choice in theorizing about personal identity. If her justificatory standards for hypotheses about personal identity are sometimes theoretical and sometimes practical—as the two-standpoints strategy suggests—then perhaps her different conclusions are mutually consistent as long as she holds them at different times and for different reasons. She may justifiably believe *p* under certain conditions, and $\sim p$ under other conditions.¹³ But, if so, her conclusions are also apparently consistent with the practical consequences Parfit draws from his theory of personal identity (see §3 above). By provisionally accepting Parfit’s theoretical conclusions at the outset, Korsgaard unwittingly fortifies Parfit’s metaphysical theory of personal identity from her two-standpoints attack. If there exists a stark distinction between standards for rational belief and rational action, Korsgaard’s best response is that people sufficiently bothered by Parfit’s reductionist view should try to ignore it when deciding how to act.¹⁴ This, however, is

a far cry from showing that it is not an accurate depiction of personal identity, or that, even if it is, it has no practical consequences.

Korsgaard concludes her critique of Parfit the same way she begins it: by reminding us that “we think of living our lives, and even of having our experiences, as something that we *do*. And it is this important feature of our sense of our identity that Parfit’s account leaves out” (1989, p. 121). Yet Parfit leaves this out only in the content of a metaphysical description of personal identity. He understands the common conviction that we have our experiences and that we are identical to all non-branching person-stages with whom we share a body.¹⁵ But it is precisely these phenomenal convictions that Parfit holds up to scrutiny. We cannot undermine his reductionist view by reminding him what the standard view of personal identity actually is, or by informing him that, from a certain perspective or under certain conditions, we cannot help but believe it.

7. Has Hume Refuted Blackburn?

’Tis certain there is no question in philosophy more abstruse than that concerning identity. . . . We must have recourse to the most profound metaphysics to give a satisfactory answer to it.

Hume, *Treatise* I.2.6

Perhaps the most surprising philosopher to criticize Parfit’s theory of personal identity on broadly Kantian grounds is one of Hume’s greatest allies—Simon Blackburn (1997). Blackburn initially construes the problem of personal identity as the problem of reconciling the apparently inconsistent conclusions generated by thinking about persons from a first-person or practical standpoint, on the one hand, and a third-person or theoretical standpoint, on the other. Blackburn focuses on what he calls “the unity reaction,” which reflects the conviction that no matter how cases involving severe psychological discontinuity are described from the outside (or theoretically), each resulting stream of consciousness would experience itself as a unified agent from the inside. The unity reaction is a necessary feature of conscious life. Consequently, Blackburn thinks, “even if Parfit’s metaphysics survives, its ethical consequences are less than he imagines” (*ibid.*, p. 183).

Blackburn echoes Korsgaard’s emphasis on viewing problems of personal identity from the practical point of view: “what I want to propose is that a certain kind of activity, with a consequent purely formal unity of the self, is involved in any kind of practical reasoning” (*ibid.*). According to Blackburn, when we reason practically, we regard ourselves as unified agents, and this self-conception always induces us to give definitive answers one way or another to Parfit’s thought experiments. Even if our metaphysics leads us to believe that survival can come in degrees, we are under a practical “compulsion” to regard our identity as determinate, and

this compulsion “entirely survives the awareness that the empirical self will have dissolved or merged or split in the cases set up” (*ibid.*).¹⁶

Blackburn’s argument for the unity of agency over time hinges on a straightforward empirical claim. He believes it is a psychological fact about creatures like ourselves that, when thinking about the future, we regard ourselves at each stage of the future as identical. That is, at each stage along the spatio-temporal path our body traces, we necessarily consider ourselves identical to all previous and future stages. According to Blackburn, when we deliberate about the future we cannot think of ourselves as merely more or less related to the person we once were or the person we anticipate becoming.

This claim is falsifiable, and we only need a single counter-example to falsify it. Consider someone who is diagnosed with a degenerative brain disease, such as Alzheimer’s. In thinking about the distant future, he may come to accept that his future self will bear only some degree of Relation R to his present self. If this is true, Blackburn’s claim that he is under a practical compulsion to think of his identity as determinate is false.¹⁷ Upon careful reflection, he may not believe that he will simply be or not be that future person. And, although it may conflict with ordinary thought, he apparently can believe this while he deliberates on what he should do now. To take a more mundane example, it seems undeniably true that each of us loses some continuity with the child we used to be, and most of us do not believe we are psychologically identical to the child we once were.

Blackburn balks at the traditional Kantian arguments for the unity of agency over time. For example, he rejects Peter Strawson’s attempt to ground personal identity in the necessity of positing a unified self over time for the very possibility of experience. Blackburn considers it the wrong strategy because it involves a third-person, theoretical approach, which is vulnerable to the kinds of metaphysical difficulties Parfit’s science fiction cases raise. Our worries about personal identity, Blackburn thinks, are practical, rather than theoretical. The trick for defending something like the standard view of personal identity, then, is to abandon third-person theorizing altogether (when thinking about personal identity) in favour of a practical, first-person self-conception. Blackburn eloquently expresses the crux of the difference over personal identity between neo-Humeans, such as Parfit, and neo-Kantians, such as Korsgaard:

If Kant is right that the two perspectives of agency and judgement, on the one hand, and of anything found in an objective, Humean or Parfitian ontology, on the other, simply do not mix, then it will be tempting to give complete metaphysical sovereignty to the latter, and to downgrade the former as at best a question of a necessary illusion somehow required for thinking of oneself as other than passive. (Blackburn 1997, p. 191)

In contrast, Blackburn continues, “the [Kantian] tendency is to allow the reductionist metaphysics, but to try to sever it from its ethical consequences” (ibid.). Blackburn reluctantly embraces this approach. Although thinking of Kant’s argument in terms of two standpoints is better, Blackburn maintains, than thinking of it in terms of two (ontologically distinct) worlds, problems remain. How do we reconcile the two perspectives when they deliver inconsistent conclusions? Moreover, how can we explain or describe the primacy of the practical perspective without invoking the theoretical standpoint?

8. Blackburn and the Primacy of the Practical

Despite his reservations about the two-standpoints strategy, Blackburn ultimately accepts some version of it, granting primacy to the practical standpoint because it better represents our ordinary conception of practical reasoning. According to Blackburn, Parfit’s treatment of personal identity as a metaphysical theory “misrepresents” the way we think of ourselves when reasoning about what to do. Following Korsgaard, Blackburn insists that when we deliberate, we wonder what we should think or how we should act; we do not wait to see what we end up believing or how we end up acting. When reasoning practically, Blackburn says, “I necessarily think of the possibilities as open”; for instance, “in wondering whether to show my holiday slides or offer a drink, I am not wondering how events . . . will unfold. I am wondering how to unfold them” (ibid., p. 193). The implication is that we can answer any interesting question about personal identity from the practical perspective, though we can always answer trivial questions about the degrees of R that obtain between successive person-stages from the theoretical perspective (ibid., p. 195). No matter what our theoretical beliefs are, Blackburn implies, when we decide such things as whether we should save money, have children, or change careers, we are always led back to viewing ourselves from the inside. No matter what we discover about the degree to which certain psychological connections obtain between particular person-stages at different times, these “external” discoveries do not affect how we answer “internal” questions about what we ought to do.

Although this is often true, there are at least two problems with this claim. First, Blackburn’s contentions that we “necessarily” think of the possibilities as open and “necessarily” think of our identity as determinate when occupying the practical standpoint does not imply that we are correct to conceive of ourselves in this way.¹⁸ If personal identity in fact depends upon a relation (such as R+U) that we fail to acknowledge, then our failure to acknowledge it is irrelevant: it does not reveal a different, practical truth. Similarly, to continue with Korsgaard’s analogy between questions about personal identity and free will, even if I am psychologically unable to conceive of my will as causally determined when I occupy

the practical standpoint, if my will is in fact fully causally determined (whether or not I can know this fact), then the phenomenal feeling of free will is beside the point.¹⁹ My practical conviction is deeply inconsistent with the facts. In the case of personal identity, when our practical conviction in the standard view conflicts with our theoretical beliefs about the link between Relation R and personal identity, we are tempted—correctly, I think—to give primacy to the theoretical conclusion.

Unlike Blackburn, I believe this temptation reflects a deep insight rather than a philosophical prejudice. One reason to think so is that if we reject the general primacy of our theoretical conclusions, we inadvertently undermine the authority of the empirical and causal beliefs on which practical deliberation depends. When we think about how to efficiently satisfy a desire, for example, we invoke theoretical beliefs about both the existence of our goals and the causal means to their satisfaction. When I ask myself whether I ought to take the A train or the Q train to Brooklyn, I take for granted a variety of beliefs about the direction of train routes, the accuracy of subway maps, and so on. And these beliefs are not themselves the outcome of practical reasoning; they are the foundational facts (or beliefs) on the basis of which practical reasoning proceeds. This implies that practical deliberation cannot rest on firm ground until we take certain fundamental classes of theoretical beliefs for granted. Among these may be the psychological and metaphysical facts that comprise Relation R.

The second problem with Blackburn's claim that we are always led back to the standard view of personal identity when reasoning practically is that, like his previous argument about unified agency, it is an empirical description of how we respond to theoretical discoveries; as such, it is empirically defeasible. People respond to insights about personal identity differently. Wounds in our theoretical views can bleed into our practical convictions.²⁰ And this is enough to refute Blackburn's claim that theoretical insights cannot affect practical questions about personal identity in an interesting way. Parfit, for example, confesses that his metaphysical beliefs about personal identity have both decreased his concern for his distant future and removed a barrier between concern for himself and concern for others: "Thinking hard about these arguments removes the glass wall between me and others. And . . . I care less about my death [since it] is merely the fact that, after a certain time, none of the experiences that will occur will be related, in certain ways, to my present experiences" (1984, p. 282). Parfit concedes that "this is simply a report of psychological effects" and that "the effects on others may be different" (*ibid.*).

But there are also ways in which our theoretical insights must (insofar as we are rational) impinge upon our practical beliefs. If we initially hold the standard view of personal identity, but then come to accept a psychological criterion and with it come to believe that Teletransportation is a

form of travel rather than suicide, then when we reason practically about whether to take an airplane to Australia or a Teletransporter, we are *ceteris paribus* irrational to choose the plane.²¹ Blackburn cannot simply appeal to the practical standpoint in deciding whether to fly or Teletransport. Theoretical beliefs may very well bear upon our choice about what to do and how to conceptualize what each alternative involves. Thus, for example, if we are concerned to minimize the amount of pain in the world, and we discover that animals with sophisticated nervous systems are not the automatons Descartes thought they were, but are instead sentient creatures, this must (insofar as we are rational) affect how we treat them. In other cases, theoretical discoveries, including alterations in our metaphysical beliefs, delineate a range of rationally defensible practical beliefs.

As mentioned earlier, Parfit thinks one way we might respond to the realization that survival depends upon the degree to which Relation R obtains between successive person-stages is to adjust our concern for our distant future in certain ways. We might have a discount rate not merely with respect to time, but “with respect to the weakening of one of the two relations [psychological continuity] which are fundamentally what matter [in questions about survival]” (1984, p. 314). Parfit does not think the reductionist view entails this attitude toward future person-stages, but that it rationally permits it, and that—for some people—the reductionist view may help support this attitude.

Parfit’s example of a nineteenth-century Russian socialist provides a vivid illustration. Parfit imagines an idealistic young man who anticipates inheriting valuable property in the distant future. His current intention is to distribute that land to peasants when he inherits it. But he believes that as he ages he may develop a stronger desire to keep the land for himself. So he decides to sign a contract to compel himself, as the person he is then, to give up the land. He then implores his wife, “I regard my ideals as essential to me. If I lose these ideals, I want you to think that I cease to exist. I want you to regard your husband then, not as me, the man who asks you for a promise, but only as his corrupted later self. Promise me that you would not do what he asks” (ibid., p. 327). The case illustrates how as people change, they may find past commitments or ideals—ideals which partly constitute the self—deeply repugnant, even alien.²² *Pace* Blackburn, the theoretical fact that your future self may become alien to your current self may justifiably guide your decision about what to do now.

9. Concluding Remarks

Rather than defending Parfit’s endorsement of utilitarianism as the inevitable outcome of his reductionist view of personal identity, I have defended the thesis that our metaphysical (or, more generally, theoretical) beliefs can rationally alter both the scope and content of our practical reasons. Korsgaard seems to think that our practical belief in what Parfit

dubs the “standard view” of personal identity is compatible with Parfit’s metaphysical theory, while Blackburn thinks we can accommodate Parfit’s metaphysical conclusions once we recognize that all interesting questions about personal identity remain practical. I have argued that although both Korsgaard and Blackburn may accurately describe how personal identity is commonly conceived, their attempts to rescue the standard view entail conclusions that we have little reason to accept. Korsgaard’s version of the two-standpoints strategy either invites contradictory beliefs about personal identity or leads to an objectionable form of relativism. Blackburn, on the other hand, takes the position that all interesting questions about personal identity are practical, rather than metaphysical, and thereby forecloses the possibility of incorporating theoretical beliefs into practical deliberation.²³

Notes

- 1 Nevertheless, Bernard Williams has rightly argued (in response to an earlier article of Parfit’s) that even if Parfit shows that the differences between particular persons over time can be significant, he does not show that the differences between discrete persons at any given time are insignificant (Williams 1976a, p. 202). And this seems to be the main worry expressed by critics of utilitarianism.
- 2 The “two-standpoints” distinction is most clearly developed in §III of Kant’s *Groundwork of the Metaphysics of Morals* (1997).
- 3 Parfit encapsulates his reductionist view of persons as follows: “the existence of a person, during any period, just consists in the existence of his brain and body, and the thinking of his thoughts, and the doing of his deeds, and the occurrence of many other physical and mental events” (1984, p. 275). Notice that a reductionist view of personal identity entails a reductionist view of persons: if personal identity over time consists in some set of facts about physical or psychological continuity, then personhood at a particular time must likewise consist in facts about physical or psychological phenomena. Although Parfit never explicitly connects the two, his reductionist view encompasses both persons and personal identity.
- 4 Parfit thinks persons are conceptually but not ontologically distinct from the relevant physical and mental activity that comprise them. He calls this sort of reduction constitutive rather than eliminative: “On Constitutive Reductionism, the fact of personal identity is distinct from these facts about physical and psychological continuity. But since it just consists in them, it is not an independent or separately obtaining fact” (Parfit 2003, p. 299).
- 5 Of course, the ways in which an intention can be fulfilled may vary. The article I now write, for example, differs somewhat in content from the article I had originally envisaged.
- 6 Unlike identity, neither component of Relation R is transitive: with memory, for example, from the fact that Picasso at age seventy-five recalls an experience

of Picasso at age fifty, and Picasso at fifty recalls an experience of Picasso at twenty-five, we cannot conclude that Picasso at seventy-five recalls an experience of Picasso at twenty-five.

- 7 Since rational choices are made on the basis of evidence rather than facts about a relevant prospect, Parfit should instead argue that it can be rational to care less about one's future self when one justifiably anticipates discontinuity between one's current self and future self. This might occur, for example, given the knowledge that one has a genetic predisposition to Alzheimer's disease and, consequently, a high probability of severe dementia in old age.
- 8 It is important to notice that Korsgaard uses "identity" here to mean personhood or personal identity in the colloquial sense of that term. A large part of her essay is concerned with the issue of precisely what a person is. However, her reference to continuity and time in the passage cited above signals that this particular argument concerns personal identity over time.
- 9 Although this supposition is presented as a concession to Parfit, Parfit actually expresses reservation about the term "successive self" because in most cases, he thinks, Relation R does not diminish so much that we are warranted in talking about several discrete selves (Parfit 1984, p. 306).
- 10 Korsgaard reaffirms this argument in *The Sources of Normativity* (1996), §3.2.2.
- 11 Nozick draws the distinction and discusses the example in *The Nature of Rationality* (1993, p. 70). Parfit makes a similar distinction (1984, p. 324).
- 12 Nozick correctly suggests that theoretical conclusions are, in general, to be ranked lexically prior to practical beliefs in terms of their overall credibility (1993, p. 87).
- 13 For instance, she may justifiably believe persons persist over indefinitely long periods of time when deciding how to treat people, given her commitment to friendship, but she may also justifiably believe that persons are a series of more or less connected selves when thinking about matters metaphysically.
- 14 Although this may be Korsgaard's best response, I doubt she would actually endorse it. Nor is it usually psychologically possible to set aside evidence that bears on one's central beliefs about the world. On this point, see Bernard Williams (1976b).
- 15 "Person-stage" is Nozick's term, not Parfit's, but it allows us to avoid such awkward locutions as "one's future self," which falsely implies identity with that self.
- 16 Notice that I have reverted to talk of "identity" rather than "survival," which is the real focus here. If $PI = R+U$, then the cases in question obviously do not involve questions of identity, but rather degrees of survival. But since Blackburn prefers to use "identity" in discussing Parfit's views, I will follow him in this section to avoid vacillating between his and Parfit's preferred usage.
- 17 Parfit hints at a similar answer in considering the difficulty involved in accepting the reductionist view: "since I can believe this view, I assume that others can do so too. We *can* believe the truth about ourselves" (1984, p. 280).

- 18 In fact, in another context, Blackburn dismisses the “Kantian fantasy” that we can transcend certain emotional and physical facts about ourselves in order to create an identity: “You, when you deliberate, are what you are: a person of tangled desires, conflicting attitudes to your parents, inchoate ambitions, preferences, and ideals, with an inherited ragbag of attitudes to different actions, situations, and characters. You do not manage, ever, to stand apart from all that” (1998, p. 252).
- 19 Henry Sidgwick made a similar point in response to Kant’s “postulates” of practical reason: “the mere fact that I cannot act rationally without assuming a certain proposition, does not appear to me—as it does to some minds—a sufficient ground for believing it to be true” (1981, p. 472).
- 20 Obvious examples include how our beliefs about free will impact our theories of desserts and punishment, and how our beliefs about the (non-)existence of an afterlife impact our beliefs about how we ought to live our lives.
- 21 Assuming, of course, that Teletransportation is reliable, cheap, etc.
- 22 It might be thought that Parfit’s emphasis on ideals regarded as constitutive of oneself vindicates Blackburn’s belief that questions about personal identity must be answered from a first-person, practical standpoint. However, Parfit can respond as follows. We each define ourselves at any given point in terms of certain distinctive experiences, desires, and character traits. But once this is done, the distinctive psychological features can be understood as an external, theoretical fact about that person, which can be measured for continuity over time. Thomas Nagel accordingly argues that although objective or “external” truths about mental life must originate from a subjective or “internal” perspective, they need not stay there: “What is more subjective is not necessarily more private. . . . The subjective ideas of experience, of action, and of the self are in some sense public or common property” (1979, p. 207).
- 23 The author would like to thank Margo Adler, Oliver Sensen, Graham Forbes, and two anonymous referees from this journal for their helpful comments on an earlier draft.

References

Blackburn, Simon

1997 “Has Kant Refuted Parfit?” In *Reading Parfit*. Edited by Jonathan Dancy. Oxford: Blackwell Publishing, pp. 180-201.

1998 *Ruling Passions*. Oxford: Oxford University Press.

Hume, David

2000 *A Treatise of Human Nature*. Oxford: Oxford University Press.

Kant, Immanuel

1997 *Groundwork of the Metaphysics of Morals*. Translated by Mary Gregor. Cambridge: Cambridge University Press.

Korsgaard, Christine

1989 "Personal Identity and the Unity of Agency: A Kantian Response to Parfit." *Philosophy and Public Affairs*, 18: 101-32.

1996 *The Sources of Normativity*. New York: Cambridge University Press.

Nagel, Thomas

1979 "Subjective and Objective." In *Mortal Questions*. Cambridge, MA: Cambridge University Press, pp. 196-214.

Nozick, Robert

1993 *The Nature of Rationality*. Princeton, NJ: Princeton University Press.

Parfit, Derek

1973 "Later Selves and Moral Principles." In *Philosophy and Personal Relations: An Anglo-French Study*. Edited by A. Montefiore. Montreal and Kingston: McGill-Queen's University Press, pp. 137-69.

1984 *Reasons and Persons*. Oxford: Oxford University Press.

2003 "The Unimportance of Identity." In *Personal Identity*. Edited by Raymond Martin and John Barnes. Oxford: Blackwell Publishing, pp. 292-317.

Rawls, John

1999 *A Theory of Justice*. Cambridge: Harvard University Press.

Scheffler, Samuel

1982 "Ethics, Personal Identity, and Ideals of the Person." *Canadian Journal of Philosophy*, 12: 229-46.

Sidgwick, Henry

1981 *The Methods of Ethics*. Indianapolis, IN: Hackett.

Williams, Bernard

1976a "Persons, Character, and Morality." In *The Identities of Persons*. Berkeley, CA: University of California Press, pp. 197-216.

1976b "Deciding to Believe." In *Problems of the Self*. Cambridge: Cambridge University Press, pp. 136-51.

Wolf, Susan

1986 "Self-Interest and Interest in Selves." *Ethics*, 96: 704-20.