

The Personal Is Political

by

Jeff Sebo

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Philosophy

New York University

September, 2011

J. David Velleman

Acknowledgements

For helpful comments on drafts of these papers, I am grateful to David Chalmers, Jonny Cottrell, Justin Clarke-Doane, Andy Egan, Richard Galvin, Blake Hestir, Kelly Heuer, Dale Jamieson, Matt Kotzen, Colin Marshall, John Morrison, Thomas Nagel, Nathana O'Brien, Japa Pallikkathayil, Mike Raven, Gina Rini, William Ruddick, Sam Scheffler, Jon Simon, and the participants at the NYU Thesis Preparation Seminar, the 2nd Rocky Mountain Ethics Congress, and the University of Washington Graduate Student Conference on Moral Psychology. I am especially grateful to Derek Parfit, John Richardson, Sharon Street, and J. David Velleman for serving on my dissertation committee, the latter as chair. All four of my advisers had a profound impact on the development of my philosophical thought, which is exactly what I hoped for when I asked them to work with me.

Finally, thanks to my mom, dad, and brother for being my mom, dad, and brother; thanks to Maryse Mitchell-Brody for being my lady; and thanks to my night self for inspiring many of my ideas here, as well as for leaving me with enough time and energy to write at least some of them down.

Contents

- Acknowledgements ii**

- Introduction 1**

- 1. The Subject As Multiplicity6**
 - 1.1. Three Kinds of Psychological Fragmentation 8**
 - 1.2. The Costs of Psychological Fragmentation 24**
 - 1.3. The Benefits of Psychological Fragmentation 29**
 - 1.4. Practical Integrity..... 51**

- 2. Is Self-Binding Morally Wrong?..... 56**
 - 2.1. Fusion 59**
 - 2.2. Multiple Personalities 92**
 - 2.3. The Just Soul 98**
 - 2.4. Conclusion 107**

- 3. Personal Responsibility 109**
 - 3.1. Different Senses of ‘Responsibility’ 113**
 - 3.2. Personal Responsibility 135**
 - 3.3. Conclusion 153**

- Conclusion 155**

- Works Cited 164**

Introduction

Plato argued in *The Republic* that, since the soul is divided, we should use political morality as a model for personal morality. I agree with this, though I disagree about many of the details. Specifically, Plato claimed that the soul is divided into a rational part, a spirited part, and an appetitive part. In contrast, I believe that many people have *multiple personalities*, in the non-pathological sense that we tend to have at least some different beliefs, desires, aims, habits, and so on in different contexts. Moreover, Plato claimed that the just soul is a *dictatorship*, where the rational part rules over the spirited and appetitive parts. In contrast, I believe that the just soul is a *liberal democracy*, where, if we have multiple personalities (even in a non-pathological sense), then our personalities have a duty to jointly commit to a fair set of laws, so that they can (a) get along with each other and (b) get along, as a person, with other people.

In this dissertation, *The Personal Is Political*, I develop this analogy with the state in relation to three moral issues: integrity, justice, and responsibility.

First, in Chapter 1, “The Subject As Multiplicity,” I apply this analogy to the question of psychological self-integration. I begin by considering three ways in which a person can be “fragmented”: we can have different *personas*, we can have different *personalities* (in this non-pathological sense), and we can have different *self-conceptions*. I then argue that these kinds of fragmentation have many of the same

costs and benefits for people that they do for groups of people like states: they cause disagreement and conflict, but they also allow us to perform experiments in living, adopt a division of labor, compartmentalize harm, and learn from each other and grow more tolerant and open-minded as a result. Thus, just as many groups of people ultimately benefit from a balance between unity and diversity, many people ultimately benefit from a balance between unity and diversity as well.

Next, in Chapter 2, “Is Self-Binding Morally Wrong?,” I apply the analogy with the state to the question of *justice*, as well as to the related question of *sovereignty*. Specifically, I argue that personalities who live in the same body have the same kind of moral relationship with each other as people who live in the same nation: they disagree about how to live, yet they also have to live *together*. Thus, they have a duty to jointly commit to a fair set of laws, as well as a right to bind each other only insofar as they *have* to, in order to enforce these laws. Moreover, they have this duty even if they share a chain of memory, a high degree of psychological connectedness, and a self-conception. I then consider the political implications of this picture of intrapersonal morality. In particular, I argue that a person with multiple personalities has a “just soul” insofar as, and only insofar as, their personalities jointly commit to a fair set of laws. And a person with multiple personalities is sovereign, i.e. has the right to live the way they want, insofar as, and only insofar as, their soul is just.

Next, in Chapter 3, “Personal Responsibility,” I apply the analogy with the

state to the question of *responsibility*. Specifically, I claim that, if my arguments in chapter 2 are correct, then each personality is, at best, praise- and blameworthy only for what it does. This might seem like an implausible claim. But I argue that, even if each personality is praise- and blameworthy only for what it does, they can still be ‘morally responsible’ for what each other does in many ways. For example, they can be *collectively* praise- or blameworthy for what they do together, and they can also be indirectly responsible for what each other does, criticizable in light of what each other does, and liable for what each other does. Moreover, I argue, if personalities share a body/brain, a chain of memory, a high degree of psychological connectedness, and a self-conception, then they are likely to be morally responsible for everything that each other does in all, or at least most, of these senses very often. Granted, they might not always be. For example, if you take a trip to Vegas, and this trip brings out a new personality that decides to have an affair – an action that your other personalities never planned, could never have foreseen, would never have performed themselves, and so on – then my conception of agency implies that your other personalities are not blameworthy for this action when they get home, but are rather only somewhat criticizable in light of it and liable for it. But I argue that, in these cases, this is a plausible result.

Finally, in a brief conclusion, I consider a question that my arguments in this dissertation raise about the nature of agency. The question is: what theory of agency could imply that my “day self” and “night self” are separate agents? I canvass several

theories that might explain this result – a personality theory of agency, a revised psychological connectedness theory of agency, and a revised self-conceptions theory of agency – before sketching the alternative that seems most plausible to me, which I call the temporal self theory of agency. On this theory, each temporal self is an agent, and thus a personality is like a *political party*: it is a group of agents who have a lot in common, and who are therefore especially likely to share intentions with respect to many issues – though they might not always share intentions with each other, and might sometimes share intentions with members of other groups. In my view, the temporal self theory, while radical in some respects, ultimately provides the best explanation of my conclusions in this dissertation, while also avoiding all the problems that the other theories that I consider face.

In any event, whatever we decide about the nature of agency, we can at the very least draw the following conclusion from my arguments in this dissertation: there is no difference in kind between interpersonal morality and intrapersonal morality insofar as it applies to people with multiple personalities. Rather, there is only a difference in degree. If a person has multiple personalities, then their personalities have the same kind of moral relationship with each other that we have with other people. However, their personalities also share a body/brain, a chain of memory, a high degree of psychological connectedness, and a self-conception, and therefore they have a much more *intimate* relationship with each other than we have with other people, with the upshot that their duties to each other are very different *in*

practice than our duties to other people are, i.e., (a) they have a duty to “integrate” with each other more than we do with other people (though not all the way), (b) they are justified in binding each other more often than we are with other people (though not all the time), and (c) they are morally responsible for what each other does more often than we are for what other people do (though not all the time).

The upshot is that we can justify many of our conventional judgments about the difference between intrapersonal and interpersonal morality, while at the same time accepting that the same principles govern both sets of relationship. Granted, we cannot justify *all* of our conventional judgments about intrapersonal morality on this picture, e.g., we cannot justify the view that we should try to be fully psychologically self-integrated, that we can permissibly bind our future selves all the time, and that we are always praise- or blameworthy for our actions. But I will argue that these conventional judgments are not, on reflection, plausible anyway. Thus, the picture of morality that I develop here is not only simpler than traditional pictures (in that it dispenses with, or at least whittles away at, the need for separate principles at the interpersonal and the intrapersonal levels), but also, I will argue, more *plausible* too.

1. The Subject As Multiplicity

The assumption of one single subject is perhaps unnecessary; perhaps it is just as permissible to assume a multiplicity of subjects, whose interaction and struggle is the basis of our thought and our consciousness in general? A kind of aristocracy of 'cells' in which dominion resides? ... *My hypotheses*: The subject as multiplicity.

- Nietzsche, *Will to Power* 490.

We often think, in philosophy as well as in everyday life, that a person should try to be fully "psychologically self-integrated." That is, we often think that a person should try to have a coherent set of beliefs, desires, aims, habits, and so on. This is not a claim that we often feel the need to defend, but my sense is that many of us accept it for the following kind of reason: we think, plausibly, that a person should try to have a coherent and stable set of *plans*, as well as a set of beliefs, desires, aims, habits and so on that are consistent with the constraints set by those plans. And what would be consistent with the constraints set by our plans? Presumably: beliefs, desires, aims, habits, and so on that are *coherent* with our plans as well as with each other, like the

belief that our plans are good, the desire to take the means to our ends, and so on. Otherwise, insofar as we sometimes believe that *different* plans are good, sometimes desire to take the means to *different* ends, and so on, we are likely to encounter familiar forms of “inner conflict” including ambivalence, half-heartedness, and weakness of will. Thus, we think, we should try to be fully psychologically self-integrated. In fact, some philosophers even go so far as to say that we are not “fully autonomous” unless we meet this requirement.¹

I believe, however, that, even though full psychological self-integration has many benefits, it also has many costs; and for many people, the costs outweigh the benefits, at least to a degree. Thus, I think that, for many people, at least a degree of “psychological fragmentation” is more consistent with the constraints set by our plans than full psychological integration would be.

In this paper, I will argue for this claim as follows. I will start, in section 1, by distinguishing three kinds of psychological fragmentation: we can have different *personas*, we can have different *personalities*, and we can have different *self-conceptions*.² Then, in sections 2 and 3, I will consider the costs and benefits of each kind of

¹ For example, Sarah Buss writes that coherentist accounts of autonomy assert that “an agent governs her own action if and only if her motives (the desires that move her to act) cohere with (are in harmony with) some other attitude that represents her point of view on the action.” (Buss, 2008, “Personal Autonomy,” *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/entries/personal-autonomy/>>)

² To be clear, only one of these kinds of fragmentation, having different personalities, gives us contradictory sets of beliefs, desires, aims, habits, and so on. But the other two kinds of fragmentation cause familiar forms of “inner conflict” too, and therefore all three are worth considering here.

fragmentation, and argue that, for many of us, the benefits outweigh the costs. Finally, in section 4, I will consider the implications of my argument for “self-integration” as a virtue. Specifically, I will claim that a person is “integrated” in the relevant sense insofar as, and only insofar as, they are capable of *acting* as one – independently of whether they also *think* and *feel* as one.

1.1. Three Kinds of Psychological Fragmentation

I begin by distinguishing three kinds of psychological fragmentation: we can have different *personas*, we can have different *personalities*, and we can have different *self-conceptions*. My reason for distinguishing these kinds of fragmentation is that, as I will argue below, each kind has its own costs and benefits. Thus, we need to distinguish them in order to think clearly about whether, why, and how we ought to try to be fragmented in each of these ways.

Before I begin, a qualification. I believe that many of us already *are* fragmented in each of these ways, and I will try to defend this claim somewhat below. But my main aim here is just to argue that we *can* be fragmented in each of these ways. This will be enough for the practical arguments that I want to make in sections 2-4.

Personas

First, a person can have different *personas*, in the sense that they can act like different “versions of themselves” in different social contexts. For example, I act like one version of myself around my parents, another around my brother, another around my girlfriend, another around my friends, another around my professors, another around my students, and so on. Why do I do this? In some cases, because other people tell me to. For instance, my parents told me how to act at home, at school, at church, and so on; and so I adopted a different persona in each context, in order to comply with their instructions. In other cases, I do this because other people treat me *as though* I act a certain way, and so I act this way around them in order to satisfy their expectations about me. For instance, when I was in sixth grade, everyone in my class treated me like I was cool, so I adopted a “cool” persona around them as a result. Then, about a year later, they all started treating me like an outcast, so I adopted an “outcast” persona around them as a result. And in still other cases, I do this because I *choose* to. For instance, many of my college friends were happy-go-lucky people, so I adopted a “happy-go-lucky” persona around them in order to fit in. Similarly, many of my college professors were cynical, world-weary people, so I adopted a “cynical and world-weary” persona around them in order to fit in. And so on. (Of course, in many cases, of course, we might adopt a given persona this for all these reasons at once.)

In this sense, to use the dramaturgical metaphor made famous by Shakespeare³ and later developed by Erving Goffman and David Velleman, among others, many of us are “improvisational actors” who play different “characters” in everyday life.⁴ In many cases, these characters are all more or less grounded in reality. In my case, for example, all of the different “versions of myself” that I play are based on character traits that I really do have, at least to a degree: I really do have a cool side to my character, an outcast side to my character, a happy-go-lucky side to my character, a cynical side to my character, and so on. In other cases, however, our characters might be much less grounded in reality. For instance, people who work in the corporate world might have to act like an especially “professional” version of themselves at the office. Likewise, people who work in politics might have to act like an especially “electable” version of themselves on the campaign trail – which might mean publicly endorsing many beliefs and values that they privately want no part of. Or take an even more extreme example: many people have to act like not only

³ As Shakespeare wrote in *As You Like It*:

All the world's a stage
And all the men and women merely players:
They have their exits and their entrances;
And one man in his time plays many parts

(Thanks to Derek Parfit for this reference.)

⁴ See Goffman, 1959, *The Presentation of Self In Everyday Life* and Velleman, 2009, *How We Get Along*. Like Velleman, I believe that many of us are improvisational actors playing ourselves. But like Goffman, I believe that we play different *versions* of ourselves in different contexts. Consequently, we eventually *become* different versions of ourselves in different contexts as well, i.e., we develop a different personality for each role that we play. Thus, rather than think of ourselves as an actor who plays many different characters, as Goffman and Velleman do, I believe we should think of ourselves as a *team* of actors playing many different characters.

different versions of themselves, but also different people entirely. This is true, for example, of actors, undercover cops, undercover activists, and anyone else who, for whatever reason (good or bad), finds themselves living a “dual life”.⁵

Fortunately, as Goffman and Velleman discuss in detail, we can use many of the techniques of theatrical performance in order to play our different characters well. For example, we can have a “backstage” area where we prepare for our performances, as well as a different “stage” area for each one.⁶ We can use costumes, props, sets, and other such items to enhance each performance, as well as to further compartmentalize it from the others.⁷ We can divide other people into “teammates,” “audience members,” and “outsiders,” so that we know what role each person will be playing during each performance.⁸ And perhaps most importantly (though there are many other techniques where these come from), we can use method acting to make each performance more authentic. As Velleman describes it:

⁵ Goffman discusses several examples of these “discrepant roles,” including “informers, shills, and spotters,” on 141-166.

⁶ We have a backstage area so we can rest, practice, help our teammates, and generally speaking hide all the “dirty work” that goes into our performance from our audience. (Goffman, p. 44, 112-14). And we have a different stage area for each performance in part so that we can have “audience segregation,” which allows us to convince each audience that our relationship with them is “special and unique” (p. 49).

⁷ Goffman uses the term ‘front’ to refer to these aspects of social performance. He defines ‘front’ as “the expressive equipment of a standard kind intentionally or unwittingly employed by the individual during his performance.” The main examples he discusses are setting, appearance, and manner. p. 22-34.

⁸ For a discussion of teammates and audience members, see Goffman p. 78-105; and for a discussion of outsiders, see p. 106-140.

The method actor portrays motives, thoughts, and feelings by actually having and manifesting them, while also modulating their manifestations to fit the part that he is playing. ... He channels the outpourings of his emotion into his conception of how the character would behave, relying on the emotion to carry him through his enactment of the role.⁹

If we use all these techniques, then we will be able to play many characters in everyday life as authentically as possible (though this is not to say that our performances will be fully authentic¹⁰).

Personalities

Second, we can have multiple *personalities*, in the sense that we can be disposed to have at least some different beliefs, desires, aims, habits, and so on in different contexts.

⁹ Velleman, 2009, p. 12. For more on method acting, improvisational theater, and everyday interaction, see Lee Strasberg, *A Dream of Passion: The Development of the Method* (New York: Plume, 1988) and R. Keith Sawyer, *Creating Conversations: Improvisation in Everyday Discourse* (Cresskill, NJ: Hampton Press, 2001) (as cited in Velleman 2009).

¹⁰ As Velleman points out, this method “still involves duplicity ... because [the actor] has deliberately summoned up his emotion with the help of thoughts about matters unrelated to the drama (such as the thoughts that Strasberg called ‘sense memory’)” (12)

Of course, when most people think of multiple personalities, they think of “personality disorders” like multiple personality disorder, borderline personality disorder, or addictive personality disorder.¹¹ And this is for good reason: these are especially clear examples of multiple personality. For instance, if a person has multiple personality disorder, then their personalities are fully “compartmentalized” and therefore may have very little in common: they may have a “guardian” personality who takes care of everyone, a “renegade” personality who lashes out at everyone, a “child” personality who hides from everyone, and so on. In contrast, if a person has borderline personality disorder, then their personalities are likely to be integrated to some degree, yet still fragmented enough that the transition from one to the next is clear, even jarring: for instance, when they switch from their “normal” personality to their “manic” personality, or from their “manic” personality to their “depressive” personality, everything about them, from their physical appearance to their way of interacting with people, changes. Similarly (and, of course, there are many other examples of personality disorders where these came from), if a person has addictive personality disorder, then their personalities are likely to be integrated to some degree, yet still fragmented enough to be at war over the relevant addictive behavior: for example, their “normal” personality might do everything they can to

¹¹ I am following convention in using the term ‘disorder’. However, I want to leave it open whether these disorders are always all-things-considered bad for the people who have them.

get their “addict” personality to stop eating, drinking, smoking, gambling, and so on, whereas their “addict” personality might do everything it can to continue.

But even though these personality disorders, and others like them, are especially clear examples of multiple personality, many psychologists now believe that we can have multiple personalities in a less apparent, and less pathological, sense as well – and this is the sense that I primarily have in mind here.¹² Specifically, we

¹² Many psychologists and philosophers have written about the fact that our personalities seem to change from situation to situation. But they disagree about what the upshot is for personality psychology. *Dispositionists* say that we have internal character traits, but some of these character traits are more apparent in some situations than in others. In contrast, *situationists* say that we have no internal character traits at all; thus, our behavior is caused by external features of the environment. Many psychologists, however, now believe that the truth is somewhere in the middle: many of us have different, but overlapping, *sets* of internal character traits, and each set comes out in a different context. Paul Bloom describes this view as follows:

It is conservative in that it accepts that brains give rise to selves that last over time, plan for the future, and so on. But it is radical in that it gives up the idea that there is just one self per head. The idea is that instead, within each brain, different selves are continually popping in and out of existence. They have different desires, and they fight for control – bargaining with, deceiving, and plotting against one another. (Bloom, Paul, 2008, “First-Person Plural.” *The Atlantic*.)

I believe that this middle-ground, “interactionist” view is plausible, and in what follows, I will assume that it is true – or at least that it *can* be true, for some of us. (That said, I will not take a stand here on whether having multiple personalities, in this sense, amounts to having multiple *selves*.)

For philosophical discussion of the idea that our personalities change from situation to situation, see John Doris, 2005, *Lack of Character: Personality and Moral Behavior*, Cambridge: Cambridge University Press, and Gilbert Harman, 2000. “The Nonexistence of Character Traits.” *Proceedings of the Aristotelian Society* 100: 223-226. And for philosophical discussion of multiple personalities (in a pathological and non-pathological sense), see Ian Hacking, 1998, *Rewriting the Soul: Multiple Personalities and the Science of Memory*, Princeton: Princeton University Press.

can have multiple personalities that are integrated to a *very* high degree – enough so that the transition from one to the next is nearly seamless, even from the inside – yet still fragmented enough that they have at least some different beliefs, desires, aims, habits, and other character traits. For example, whenever I spend time with my college friends, I not only *act* more happy-go-lucky than I usually do but also *feel* more happy-go-lucky than I usually do. Similarly, whenever I spend time with my college professors, I not only *act* more cynical and world-weary than I usually do but also *feel* more cynical and world-weary than I usually do.

Of course, this is not an accident. If we play different characters in everyday life, and if we use all the techniques of theatrical performance in playing these characters, then this can cause us to *become* each of the characters that we play, at least to a degree.¹³ For example, my having a “backstage” area where I prepare for my performances as well as a different “stage” area for each performance allows me to compartmentalize each performance, as well as to “disappear” into each character as a result. My using different clothes, props, and sets for my performances allows me to further compartmentalize each performance, as well as fully disappear into each character as a result. My labeling different people as “teammates,” “audience

¹³ Carter makes the same point when she says “there is no sharp division between being a character and acting it,” and, “if you are totally immersed in a part, your thoughts, perceptions and feelings *become* those of that character” (15). As I will explain in a moment, I think this is too strong: instead of saying that we become the characters we play, we should say that we become a team of *actors*, each of whom has a lot (but not everything) in common with the character they play.

members,” and “outsiders” for each performance helps here too, because it makes it easier for me to play one role instead of several in situations where “worlds collide.” Finally, and most importantly, my approaching each role as a method actor allows me to compartmentalize each performance and disappear into each character in a very real way. In particular, if I draw upon the thoughts and feelings that I take each character to have during my performance, then, over time, the effects of psychological compartmentalization will kick in, and I will start to (a) associate those thoughts and feelings with each other, as well as with the character I draw upon them for, as well as (b) dissociate them from the thoughts and feelings that I draw upon in other contexts, while playing other characters.¹⁴

The upshot is that, in acting like different versions of ourselves in everyday life, we can experience a kind of psychological fission: we can develop a different personality for each character that we play, such that, when we enter the relevant context, the relevant personality comes out naturally. Granted, this is unlikely to

¹⁴ There are many well-known examples of method actors becoming the characters they play, and then having trouble “returning to normal”. Martin Sheen had this experience in *Apocalypse Now*, Jim Carrey had it in *Man on the Moon*, Daniel Day-Lewis has it in pretty much every movie he stars in, and so on. Rita Carter discusses many other examples too, including this one: “The British actor David Suchet ... had a long stage run in *Timon of Athens*, during which he found it increasingly difficult to flip back into [himself] when the nightly performance ended. One evening a psychiatrist friend visited him backstage and observed that he seemed still to be acting like Timon. Suchet dismissed his concern, at which point the psychiatrist shot at him a number of questions such as: ages of your children? phone number? Date of birth? To his own consternation, Suchet found he had to work hard to retrieve the answers – the Timon personality he had created was so firmly in charge that his [other] memories were temporarily irretrievable” (Carter, p. 108).

make us *fully become* any of the characters that we play in everyday life: no matter how much we “disappear into” a character, we will probably still have to *play* this character to a degree as well. For example, even though I feel more happy-go-lucky around my college friends than I do around other people, there is still a remainder, i.e., still a degree to which I *act* more happy-go-lucky around my college friends than I really am. Similarly, even though I feel more cynical and world-weary around my college professors than I do around other people, there is still a remainder, i.e., still a degree to which I act more cynical and world-weary around them than I really am. Nevertheless, these techniques can at least collapse the distinction between actor and character to a degree: rather than being an improvisational actor who plays many characters, we can be a *team* of improvisational actors, each of whom plays a character he has a lot in common with.¹⁵

¹⁵ We can also have the same kind of experience when we *imagine* acting like other people. For example, when we read books, watch movies, play videogames, play cops-and-robbers in the backyard, and so on, we might spend so much time with particular characters, or even just in particular worlds, that our personality changes a bit as a result—for the duration of the experience, and maybe for a while after. Paul Bloom describes this experience well:

Enjoying fiction requires a shift in selfhood. You give up your own identity and try on the identities of other people, adopting their perspectives so as to share their experiences. This allows us to enjoy fictional events that would shock and sadden us in real life. When Tony Soprano kills someone, you respond differently than you would to a real murder; you accept and adopt some of the moral premises of the Soprano universe. You become, if just for a moment, Tony Soprano. (2008)

Further, the experience of “getting lost” in a fictional world, and developing new personalities as a result, can be even more pronounced for the *creators* of these

Self-Conceptions

Finally, and relatedly, we can have multiple *self-conceptions*, in the sense that we can have different conceptions of who we are and what we care about. Specifically, if we have different personas and different personalities, then we can have (a) a conception of each of our *personas*, i.e., a conception of who each character is and what they care about, (b) a conception of each of our *personalities*, i.e., a conception of who we are and what we care about *within* each context, and (c) a conception of ourselves as a *person*, i.e., a “simplified” or “idealized” conception of who we are and what we care about *across* contexts. (The reason these all count as self-conceptions, despite being conceptions of different things, is that they all play the same kind of role in our practical lives: they all attribute a particular set of beliefs, desires, aims, and other character traits to us, and we can identify with them all in everyday life when deciding how to act.)

fictional worlds. In fact, some authors report that the characters they create eventually become so “real” that they tell the writers what they should say and do, and sometimes they even try to wrestle control of the story away from the writers entirely.

For more on this phenomenon, see Marjorie Taylor and A. Kohanyi, “The Illusion of Independent Agency: Do Adult Fiction Writers Experience Their Characters as Having Minds of Their Own?” *Imagination, Cognition and Personality*, 2002/3, No. 22, pp. 361-80.)

Take each kind of self-conception in turn. First, we can have a conception of each of our *personas*. These are conceptions of the characters that we play in everyday life. For example, I might have a self-conception for the happy-go-lucky character that I play around my college friends, a self-conception for the cynical character that I play around my college professors, and so on. If I do, then I can regard these as *normative* self-conceptions, in that I can use them to think about who I *ought to be* in the relevant context (or at least who I ought to *appear* to be), as opposed to who I really *am* in that context.

Second, we can have a conception of each of our *personalities*. These are conceptions of who we are and what we care about *within* each context. For example, I might think of myself as having a happy-go-lucky personality who comes out around my college friends (and who plays an *especially* happy-go-lucky character in that context), as well as a cynical and world-weary personality who comes out around my college professors (and who plays an *especially* cynical and world-weary character around them). If so, then I can regard these as *descriptive* self-conceptions, in that I can use them to think about who I really *am* in the relevant context, as opposed to who I ought to be (or at least appear to be).

Finally, we can have a conception of who we are as a *person*. This is a conception of who we are and what we care about *across* contexts. What does it mean to have a conception of who we are and what we care about *across* contexts, over and above a conception of who we are and what we care about *within* contexts? The

general idea is to construct a “simple” or “idealized” set of thoughts, feelings, and habits that we can attribute to ourselves at all times. There are several ways we might do this; let me here just mention one. Daniel Dennett has argued that we construct our self-conceptions by finding our “center of narrative gravity” across contexts.¹⁶ Here is how I understand the metaphor. Just as our *actual* center of gravity is the location around which our weight is evenly distributed, our center of *narrative* gravity is the story about who we are, and what we think and feel, in light of which all our actions and experiences make equal sense. On this view, we might construct our *personality* self-conceptions by finding our center of narrative gravity *within* each context – by telling a story about who we are and what we think and feel that makes equal sense of our actions and experiences in that context. And then we might construct our *person* self-conception by finding our center of narrative gravity *across* contexts – by telling a story about who we are and what we think and feel that makes sense of our actions and experiences *across* contexts. Thus, for example, I might think of myself as having a happy-go-lucky self who plays an *especially* happy-go-lucky character, as having a cynical and world-weary self who plays an *especially* cynical and world-weary character – and also as being a person who, “deep down,” strikes a coherent and stable balance among these seemingly conflicting character

¹⁶ Dennett, 1986, “The Self as a Center of Narrative Gravity,” in F. Kessel, P. Cole and D. Johnson, eds, *Self and Consciousness: Multiple Perspectives*, Hillsdale, NJ: Erlbaum, 1992. 15, 275-88.

traits, i.e., as being a person who, “deep down”, likes to enjoy life as much as he can while staying mindful of the fact that God is dead and life is absurd.

If we construct our person self-conceptions in this kind of way, then we can regard them as partly normative and partly descriptive. Specifically, they will be descriptive insofar as they will tell us who we *are*, rather than who we *ought to be* (or at least who we ought to *appear* to be). But they will be normative insofar as they will be *simple, idealized* conceptions of who we are, whose main purpose is not to capture our thoughts, feelings, and habits in their full complexity, but rather to “center us” around a way of life that works equally well for all our personalities. (More on this below.)¹⁷

Are We Actually Like This?

¹⁷ In my view, part of the reason why philosophers and psychologists disagree about the nature of self-conceptions is that some of them are focusing on different kinds of self-conception. For example, Korsgaard argues that our self-conceptions, or, as she calls them, “practical identities,” are *normative*, in that they are conceptions of ourselves under which we value ourselves. And she uses social roles like mother, teacher, and so on as examples of what she has in mind. (See Korsgaard, 1996, *The Sources of Normativity*, Cambridge: Cambridge University Press.) In contrast, Dennett, Velleman, and others argue that our self-conceptions are *descriptive*, in that they are conceptions of ourselves in light of which our actions and experiences make sense. (See Dennett, 1986, and Velleman, 2009.) Even Goffman, who is especially good about distinguishing actor from character, claims, for reasons that are not at all clear, that our character self-conceptions represent our “truer” selves. (p. 19, 252-3) But I think that we can resolve these disagreements just by acknowledging that we can, and many of us do, have multiple self-conceptions, some of which are normative, some of which are descriptive, and some of which are both.

My goal in distinguishing these kinds of psychological fragmentation has been to argue that many of us *can* be fragmented in these ways. Nevertheless, I think it is worth mentioning that, in the view of many personality psychologists, many of us already *are* fragmented in each of these ways. We play different versions of ourselves in different contexts, we have different personalities in different contexts, and we describe these changes by saying that we have different “selves” or “sides” that come out around some people – where sometimes this refers to our personas and other times it refers to our personalities.¹⁸

Granted, it can be easy for us to miss, or at least underestimate, just how much some of us change in everyday life, given how seamlessly we switch from persona to persona, personality to personality, and self-conception to self-conception – and also given how much of a premium we all place on psychological coherence and stability, both factually and normatively. But if we pay attention, we can see examples of psychological fragmentation all around us. Rita Carter describes many from her own experience, including:

The boss I once endured who ran his department with cold,
impersonal efficiency but turned into a lurching, sentimental moron
after his end-of-day dose of alcohol. The so-together female colleague

¹⁸ For evidence that some people have multiple self-conceptions or self-narratives in this sense, see John Altrocchi, “Individual Differences in Pluralism in Self-Structure,” in John Rowan and Mick Cooper (eds), *The Plural Self*, Sage Publications, London, 1999. And for a detailed explanation of why people create simple, idealized person self-conceptions, see Carter, p. 26-32.

who fell into inarticulate giggles whenever a man paid her a compliment. The girlfriend who always ran any show – except in her own home, where she turned into a doormat for her vile husband. The devoted family man who turned out to have a twice-a-week gay cruising habit. The shy-as-a-mouse academic who when asked to deliver a quick talk on her area of expertise ended up hogging the microphone for two hours. (Carter, xiii)

I expect that many of us can find similar examples in our own experience. We can find especially clear examples when we consider cases where an old personality comes back out after many years of being dormant. For instance, you might go back home to see your parents or attend a class reunion, only to find yourself reverting back into the petulant teenager you used to be decades ago. Or you might see an old lover again for the first time in years, only to find that all the old feelings that you worked so hard to leave behind come flooding back. This is of course why many of us engage in practices like “self-binding”: we stay away from a certain person or place, or we take measures to restrain our behavior with a certain person or in a certain place, because we know that this person or place is likely to bring out a certain “side of us,” and we worry about what this side of us will say or do.

Finally, we can also see clear examples of psychological fragmentation in popular fiction. For example, seemingly every major modern television show is based on a premise of psychological fragmentation: on characters who play different,

conflicting roles and who develop different, conflicting personalities and self-conceptions in order to play them; or, perhaps, on characters who already have different, conflicting personalities and self-conceptions and who play different, conflicting roles in order to deal with this. Thus, for example, we see people who are spouses, parents, and friends, on one hand, and doctors, lawyers, police officers, federal agents, presidents, mob bosses, drug lords, or even serial killers, on the other. This makes for compelling drama for a number of reasons; but in my mind, one of the main reasons is that it shows us (perhaps an extreme version of) the kind of psychological fragmentation that we all experience, either in ourselves or in other people we know.

But, as I said before, my main goal here has not been to show that we are fragmented in these ways, but rather just that we *can* be. In what follows, I will consider the costs and benefits of each kind of fragmentation, and argue that, for many of us, the benefits outweigh the costs – even if it means welcoming some degree of “inner conflict” into our lives.

1.2. The Costs of Psychological Fragmentation

Start with the costs of each kind of psychological fragmentation. These are widely discussed in the literature on agency and autonomy, so we can move quickly here.

First, if we play different characters in everyday life, and if these characters have different thoughts, feelings, intentions, and so on, then we might encounter situations that make it impossible for us to play all of our characters well. This is true in several respects. First, we might play characters who prefer to spend their time and energy in conflicting ways. For example, I might present myself to my family as someone who values family more than anything (while valuing many other things too), whereas I might present myself to my students and colleagues as someone who values “making a difference in the world” more than anything at all (while valuing many other things too).¹⁹ Now suppose that I get a job offer that would allow me to make much more of a difference, but which would also require me to spend much more time at the office, and away from my family. In this case, any decision I make will require me to sacrifice my performance in one of these roles, for the sake of my performance in the other. Second, we might play characters who say and do conflicting things in everyday life. For example, suppose that I find myself at the bar with my college friends and professors at the same time. In this case, I might have to play my happy-go-lucky character and my cynical and world-weary character at the same time. And this is of course very hard, maybe even impossible, to do. For example, suppose that a Bon Jovi song comes on at the bar. My happy-go-lucky

¹⁹ This is a common problem in social role playing. Specifically, as Goffman discusses, we tend to “foster the impression that the routine [we] are presently performing is [our] only routine or at least [our] most essential one” (p. 48). The upshot is that we end up playing versions of ourselves that have different, and conflicting, priorities in life.

character would sing along with it, and my cynical and world-weary character would make fun of it. Thus, if my worlds collide in this way, I might find that I have no coherent role to play at all, i.e., no way to act that will satisfy all my audiences at once.²⁰ (As I said before, we can minimize the chance of this happening by keeping our worlds from colliding in the first place, as well as by arranging for some people to be “teammates” and other people to be “audience members” when they do. But this will not always be possible – especially when we find ourselves interacting with different friends, all of whom expect us to be authentic in all our private relationships.)

Second, if we have multiple personalities, and if these personalities have different beliefs, desires, aims, habits, and so on, then this will have epistemic as well as practical costs. Epistemically, it will ensure that some of our beliefs are false, and may be irrational. And practically, it will cause a more severe version of the same problems that having multiple personas causes. First, we might have different preferences, across contexts, about how to spend our time and energy. As a result, we might have trouble committing to a given way of life, because we might change our minds about what to do across contexts. For instance, I might, at work, feel inclined to take the new job, but then, at home, feel inclined to turn it down. And even if we *do* commit to a given way of life, we might not always feel wholehearted about it

²⁰ As Goffman puts it: “When individuals witness a show that was not meant for them, they may ... become disillusioned about this show as well as about the show that was meant for them.” p. 136

(since some of our personalities will prefer a different way of life all things considered), and therefore we will be likely to experience anxiety, doubt, regret, and weakness of will in different contexts.²¹ For example, suppose that I have *three* job options: one that would be best for my family, one that would be best for “making a difference,” and one which splits a difference between the two. Now suppose that, thinking of myself as a person who, “deep down,” values family and making a difference in equal measure, I take the third job. Since I have the same person self-conception across contexts, I might be able to stick to this plan. Nevertheless, since my family personality values family above all else, I might find myself frustrated, at home, that I took the third job instead of the first; and since my work personality values making a difference above all else, I might find myself frustrated, at work, that I took the third job instead of the second. As a result, I might never be fully satisfied with my choice, even though I always regard it as the right choice for me to have made. And this sense of frustration might make it harder for me to give my best performance at home as well as at work, no matter how hard I try.

Moreover, if we have multiple personalities, then we might encounter situations where our beliefs, desires, aims, habits, and so on conflict at the time. For example, if I meet my college friends and professors for a drink, then I might not only have to play different *characters* at the same time, but also have different

²¹ Harry Frankfurt discusses the importance of wholeheartedness in many papers, including “Identity and Wholeheartedness,” in Ferdinand David Schoeman (ed.), *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*. Cambridge University Press, 1987.

personalities come out at the same time – with the upshot that I have a conflicting set of beliefs, desires, aims, habits, and so on, right here and now. Thus, for example, if a Bon Jovi song comes on the jukebox, I might feel partly inclined to sing along with it as well as partly inclined to make fun of it, and both inclinations might be strong enough that neither course of action feels authentic at the time. Thus, if my worlds collide, I might not only have no coherent *character* to play, but also no coherent *personality* for playing it, which makes it even harder to play any given role well.

Finally, if we have multiple self-conceptions – if we have multiple stories about who we are and what we care about – then we might become confused about who we really are, and might even start to lose our sense of self altogether. After all, the fact that each self-conception tells a similar kind of story about us, combined with the fact that we (often) use them all in deciding what to do, combined with the fact that we (often) use the first-person singular to refer to them all, can make it hard for us to always tell them apart. This can lead to many practical problems. First, we might have trouble deciding how to live, since we might not be able to keep track of which self-conception represents who we are, and what we think and feel, as a team. Second, we might have trouble deciding how to act in particular situations, since we might not be able to keep track of which self-conception represents who we are, and how we think and feel, as a character. Third, even if we decide how to act in particular situations, we might have trouble acting this way, since we might not be

able to keep track of which self-conception represents who we are, and how we think and feel, as an actor.²²

Finally, even if we avoid all these other problems, we will still have to spend much more mental energy keeping our stories straight than if we just had one story about ourselves to tell.²³

1.3. The Benefits of Psychological Fragmentation

These costs of psychological fragmentation might lead us to believe that we should try to be as psychologically integrated as possible: if we have a coherent and stable set of roles, beliefs, desires, aims, habits, and so on, then we can avoid all this conflict and confusion, and, as a result, we can make decisions about how to live that we

²² Moreover, even if we decide what to do and how to do it, we might wonder whether we made the right choice. Specifically, we might have an experience that in some ways resembles weakness of will – and which may even be what we sometimes use ‘weakness of will’ to refer to: we might *feel* as though we acted contrary to what we thought, at the time of action, was best, because we decided which action was best from the standpoint of one self-conception (say, our person self-conception), yet we decided which action to perform from the standpoint of another, competing self-conception (say, our personality self-conception) – without realizing that we made the switch at all.

²³ This is a big part of why Velleman thinks that practical reason favors honesty. The reason is that if we tell people lies, then we have to keep two narratives in our head – the real one and the fake one – which requires a lot more mental energy. I agree with Velleman about this, though, as I will argue below, I think that the benefits of keeping multiple narratives in our head will often (though not always) outweigh this cost.

always feel wholehearted about. But before we reach this conclusion, we have to assess the benefits of psychological fragmentation as well as the costs, and ask ourselves whether a *balance* between integration and fragmentation might be better, for some of us, than as much integration as possible. This is, of course, how we reason about integration and fragmentation at the interpersonal level. When it comes to the composition of families, colleges, businesses, and so on, we recognize that psychological fragmentation – or as we call it here, *diversity* – has many costs, but we also recognize that it has many *benefits*; and we think that, for many groups (as well as for many of the members of those groups), a balance between unity and diversity is ultimately better, all things considered, than as much unity as possible.

Given that many of us are more familiar with the benefits of diversity in the interpersonal case, my strategy in this section will be to work from the interpersonal case to the intrapersonal case. Specifically, I will consider several well-known benefits of interpersonal diversity, and then argue that these same benefits accrue in the intrapersonal case as well. The upshot will be that, for many people (as well as for many of the personalities in those people), a balance between psychological integration and fragmentation is better, all things considered, than as much integration as possible.

First, if we live in a diverse society, i.e., if we live in a society where some people have different beliefs, desires, aims, habits, and so on than others, then we can perform

what Mill calls "experiments in living."²⁴ Specifically, we can try out different ways of life, and see which works best for us collectively as well as individually. Of course, this process will also lead to some of the problems we considered above: different people will prefer different ways of life, and some people may also be conflicted about which way of life they prefer. But generally speaking, this process will allow us to gravitate away from ways of life that we all agree are bad for us, and towards ways of life that we all agree are good for us. Thus, even if we end up with a collection of subcultures that conflict, as well as with a shared culture that works better for some people than for others, we will still be better off as a society, and as individual members of that society (for the most part), than we would have been had we never allowed ourselves to experiment in the first place.

Similarly, if we let ourselves have a "diverse psychology," then we can perform what we might call "personal experiments in living."²⁵ Specifically, we can try out different ways of life, and see which works best for us collectively (as a person) as well as individually (as personalities within that person). This is especially important early in life. As a young child, for example, you might try out many different versions of yourself as a result, to see which makes sense. You might try out some of them imaginatively, by reading books, watching movies, writing stories,

²⁴ Mill, "On Liberty" (New York: Barnes and Noble Books, 2004), 59.

²⁵ Velleman makes a similar point when he writes, "Figuring out how to live is a process of trial and error in which the trials are what Mill called experiments in living." 2009, p. 83.

playing with toys, creating imaginary companions, and so on.²⁶ And you might also try some of them out in real life. For instance, you might act playful around your family, serious around your teachers, extroverted around the “cool kids,” nerdy around the “outcasts,” and so on. Then, in the process of playing all these different versions of yourself, you might also start to *become* each version of yourself to a degree. If so, then this is good: it will show you not only how you are capable of *acting* but also who you are capable of *being* – as well as how it feels to be that way, both at the time and after the fact.

As in the interpersonal case, this process might lead to some of the problems we considered above: some of the personalities you develop might prefer different ways of life than others, and some might even be conflicted about which way of life they prefer. For example, you might like being a “cool” version of yourself at the time, but not so much after the fact. Thus, you might disagree with yourself, across contexts, about which roles to continue to play, and which personalities to continue to develop. But generally speaking, this process will allow you to gravitate away from ways of life that you agree, across contexts, are bad for you, and towards ways of life that you agree, across contexts, are good for you. Thus, even if you end up with a collection of roles and personalities that conflict, as well as a shared way of life that works better for you in some contexts than in others, you are still likely to be

²⁶ As Carter discusses, imaginary companions are especially psychologically “safe” experiments in living, since “the child can try out all sorts of social behaviors vicariously, safe in the knowledge that, should the IC do anything disastrous, the child itself is safe from the consequences” (29)

better off as a person, and as individual personalities in that person (for the most part), than you would have been had you never allowed yourself to experiment in the first place.

Second, and relatedly, if we live in a diverse society, then we can have a division of labor, where each person plays a different role in a collective endeavor. Think about how this helps in the case of the state. In order for us to survive and flourish individually as well as collectively, we need different kinds of people to do different kinds of thing. We need parents, teachers, doctors, lawyers, politicians, police officers, soldiers, and so on. And of course, different kinds of people will be better at different kinds of role, depending on their psychological profile. For example, if you are a naturally compassionate person, then you might be especially well-suited to be a teacher, since you will be naturally inclined to give your students the care and support they need. In contrast, if you are a naturally *dispassionate* person, then you might be especially well-suited to be a doctor, lawyer, politician, police officer, or soldier, since you will be naturally inclined (or at least *able*) to make the hard, life or death choices these jobs require with a minimum of anguish and remorse. Thus, since we need people to play different roles in society, and since different kinds of people are better at different kinds of role, we need, or at least benefit from, having different kinds of people in society. We benefit from having cooperative and compassionate people, as well as from having competitive and dispassionate people. We benefit from

having people who trust everyone, as well as from having people who are suspicious of everyone. And so on. As before, this will lead to many of the problems we considered above, but it will also help us to accomplish much more together than we ever could individually, or if we all played the same kind of role, or had the same kind of psychological profile.

Similarly, if we allow for a diverse psychology, then we can have what we might call a "personal division of labor," where different personalities play different roles as part of a collective endeavor. The reason is the same as before: in order for many of us to survive and flourish as people, we need to play many different roles in everyday life. We need to be sons/daughters, brothers/sisters, students, friends, employees, husbands/wives, parents, and so on. And even though we will have *some* ability to choose roles that complement each other, rather than conflict with each other, we will not have unlimited ability to do this, nor will we always want to do this. After all, many of us will have many aims in life (especially if we acquire our aims by performing experiments in living, and developing different personalities with different aims as a result), and each aim will require us to play a different role. For example, you may decide, having tried out many different ways of life, that your ideal life involves having a family and friends as well as a career – in which case you will have to play all of these roles in order to live the way you want. Moreover, many other people will have just one or two main aims in life, yet they might still find that these aims require them to play many conflicting roles. For example, you may decide

that your ideal way of life just involves having a family, but you might then find that, in order to support your family, you need to have a career so you can put food on the table, friends so you can give yourself and your family a break, and so on. Either way, then, many of us (not all of us, but many of us) will need to play many, conflicting roles in order to live the way that we want.

As when we perform personal experiments in living, we might, in the process of playing these different roles, become each of the characters we play, to a degree. If so, then, once again, this is good: it will make us better able to play each role than we would otherwise be. For example, suppose you decide that your ideal way of life involves being a parent and a politician. If you had a single, “one size fits all” personality for playing these roles, then you would have a much harder time disappearing into each one. But if, instead, you condition yourself to be more cooperative and compassionate at home and more competitive and dispassionate at work (at least, behind closed doors at work), then you will be able to play each role just by being yourself, at least to a degree. Thus, since many of us need to play many, possibly conflicting roles in order to live the way that we want, and since different kinds of personality are better for different kinds of role, many of us need, or at least benefit from, having many, possibly conflicting personalities. We benefit from feeling cooperative and compassionate sometimes, as well as from feeling competitive and dispassionate other times. We benefit from trusting people sometimes as well as from being suspicious of people other times. And so on. As before, this will lead to many of

the problems we considered above: it will cause us to disagree with ourselves across contexts, and it will make it hard for us to find a way of life that works equally well for all our personalities. But it will also help us to accomplish much more than we ever could have if we just played a single role, and/or just had a single personality.²⁷

Third, and also relatedly, if we live in a diverse society, then we can compartmentalize harm and adapt to change more easily. For example, part of the reason why it helps to have a division of labor, especially a division of labor where different people work in different industries, is that, if the market shifts in a way that makes a particular industry collapse, then it will help to have people working in other, more successful industries to carry us through the transition. Likewise, part of the reason why it helps to have people with different beliefs and values is that, if the

²⁷ For example (and there is nothing at all special about this example, other than that I happened to be reading this interview while writing this paragraph), here is what parent/stand-up/writer/director/actor/editor, etc. Louis C.K. has to say about playing all his different roles:

LCK: It's a huge amount of work, and it's also a lot of stress. It's a lot of details and decisions and pressure. I have all these different roles, and everyone who interacts with me in different roles doesn't give a shit about any of the other things I need to do, nor should they. ...

AVC: Are you able to have a personal life at this point, or is it just about work?

LCK: ... I have my kids four days, three nights a week, and we have to shoot the other three and half days. ... You know what: It actually benefits both sides, because that's the only way I can actually do this. If I had this job five days a week, I wouldn't get through it. ... That's how I regenerate. I spend time with my kids, and the stuff that I do that is very huge seems very little when I'm with them. It doesn't really matter. (<http://www.avclub.com/articles/louis-ck,58516/>)

culture shifts in a way that makes a particular belief and value system difficult to sustain, it will help to have people with other, more “resilient” belief and value systems to carry us through the transition. Of course, this will cause familiar problems too: if we have different people working in different industries, then some will be more successful than others as a result. And if we have different people with different beliefs and values, then they are likely to disagree with each other, have a hard time finding a shared way of life that works equally well for all of them, and so on. But the general upshot is that we will be much more likely to survive and flourish as a society when bad things happen – and, of course, bad things will inevitably happen.

Similarly, if we have a diverse psychology, then we can compartmentalize harm and adapt to changing circumstances as a person. This is of course a familiar point for people who study personality disorders. For example, one of the main causes of multiple personality disorder is that, when we have a traumatic experience, we cope with this by creating multiple personalities – one who remembers the traumatic experience, and another (or perhaps several others) who are protected from having to do this. But we can also use our capacity for psychological fragmentation as a *preventative* measure: specifically, we can develop multiple personalities in a non-pathological sense, where each personality has at least some beliefs, desires, aims, habits, and so on that others lack, so that, if any given personality has a traumatic experience, the others will be (relatively) protected from it. For example, part of the

reason it helps to have a personal division of labor is that, if you lose your job, or if your job becomes obsolete, then it will help to have other personalities, with other talents and interests, to carry you through the transition. And if one of your friends or family members dies (or otherwise leaves you), then you might find it helpful to be able to throw yourself into other projects or relationships for a while, thereby bringing out personalities who still feel the loss of your friend or family member, of course, but maybe not as acutely as the personality who interacted with them does. Likewise, part of the reason why it helps to have different personalities with different beliefs and values is that, if you have an experience that makes a particular set of beliefs and values difficult to sustain, then it will help for you to have other personalities with other, more “resilient” belief and value systems to carry you through the transition.²⁸

As always, these kinds of psychological divisions will cause problems. If we compartmentalize traumatic experiences, then it will be harder for the personalities who are stuck remembering them to get past them. If we have different personalities with different talents and interests, then some will be happier and more successful than others. And if we have different personalities with different beliefs and values, then we will experience inner conflict and have a hard time finding a way of life that works equally well for all of our personalities. But the general upshot is that we will

²⁸ For evidence that having multiple personalities and self-conceptions helps us compartmentalize harm in this way, see P. W. Linville, “Self-Complexity as a Cognitive Buffer Against Stress-Related Illness and Depression,” *Journal of Personality and Social Psychology*, 1987, Vol. 52, pp. 663-76.

be much more likely to survive and flourish as a person when bad things happen to us – and, of course, bad things will inevitably happen to us.

Fourth, if we live in a diverse society, then we can learn from each other, and become more tolerant and open-minded as a result.²⁹ Compare: if everyone around us had the same perspective as us, then it would be easy to dismiss other perspectives without really engaging with them. And this might lead us to accept our own perspective dogmatically, without really understanding what it means, why we accept it, or how we might make it better. But if some of the people around us have *different* perspectives than us, then it will be much harder to dismiss these other perspectives, and we will be much more likely to engage with them and understand and improve upon our own perspective as a result. This is in part for pragmatic reasons: we have to *interact* with the people around us, and therefore we have to engage with them to a degree. But it is also for motivational reasons: the more we get to know someone, the more we feel inclined to take what they have to say seriously. Thus, even if we continue to disagree with them, and even if this disagreement leads to some conflict, we will still benefit from our interaction with them, because we will understand our own perspective better, improve on it to a degree, and also become more tolerant and open-minded in general – all of which are good results, practically as well as epistemically.

²⁹ This is another point that Mill makes in *On Liberty*.

Similarly, if we have a diverse psychology, then we can learn from our other personalities and become more tolerant and open-minded as a result. The reason is the same as before. Compare: if we always had the same perspective, then it would be (relatively) easy to dismiss other perspectives without really engaging with them. And this might cause us to accept our own perspective (relatively) dogmatically, without really understanding what it means, why we accept it, or how we might improve upon it. But if we have different personalities with different perspectives, then it will be much harder for us to dismiss these other perspectives, and we will be more likely to engage with them and understand and improve upon their own perspective as a result. As before, this is in part for pragmatic reasons: we have to *interact* with our fellow personalities even more than with other people, and therefore we have to engage with them even more, as well. And it is also for motivational reasons: we remember everything that our fellow personalities think and feel from the first-person perspective. Thus, we are very familiar with what they believe and value, and all the complicated reasons why. This makes it much easier for us to take what they have to say seriously. For example, I know that I benefit from having a happy-go-lucky personality as well as a cynical and world-weary personality, because the interaction between the two makes it difficult for me to accept either view dogmatically. Specifically, every time I feel inclined to think that the world is a beautiful place, and that anyone as fortunate as me who thinks otherwise is just a stick in the mud, I also, at the same time, remember and anticipate feeling inclined to

think that the world is a *terrible* place, and that anyone as fortunate as me who thinks otherwise just has his head in the sand. As a result, I always feel compelled to challenge each view from the inside as well as from the outside, and to engage with a real (as opposed to straw man) version of the other view while doing it.

As in the interpersonal case, this kind of intrapersonal interaction may not always lead our personalities to agree with each other (though, in this particular case, my happy-go-lucky and cynical and world-weary personalities have converged considerably). But even if we continue to disagree with ourselves across contexts, we will still benefit from this interaction, since we will understand each of our perspectives better, improve on them to a degree, and become more tolerant and open-minded to a degree as well – to our fellow personalities as well as, perhaps, to other people. And these are, once again, good results, practically as well as epistemically.³⁰

Finally, if we play multiple roles and have multiple personalities, then it will be important for us to have multiple self-conceptions too, so that we can navigate all of the psychological changes that we undergo. Specifically, it will be important for us to

³⁰ Carter makes a similar point when she writes, “To be in two or more ‘minds’ about what to do is generally considered to be bad because it tends to produce the discomfort of uncertainty and inner conflict. Most of us therefore try to quash all but one of our personalities. Invariably it is the quieter, less assertive [personalities] that are made to shut up, even when they might be speaking better sense than the [personality] who vanquishes them. The problem, in other words, is not that we are multi-minded but that we refuse to acknowledge our multiplicity and so use it to our advantage.” (p. 81).

have a self-conception for each of our characters, a self-conception for each of our personalities, and a simple, idealized self-conception for ourselves as a person.

I already mentioned the role that each kind of self-conception can play for us in practical deliberation. But let me go into a bit more detail about that here. First, if we play different characters, then it will help for us to have a conception of each one. As I said before, these self-conceptions will be *normative* in that they will describe us not as we are, but rather as we ought to try to appear to be. Thus, we can use these self-conceptions to guide our decisions about what to say and do in everyday life. For example, if a politician wants to do well in a town hall meeting, then it will be important for him to keep in mind the "public" persona that he should have on display. Otherwise, he might slip up and say what he actually thinks and feels, with predictably disastrous political consequences.

Second, if we have different personalities, then it will help for us to have a conception of each one. For example, if a politician wants to do well in a town hall meeting, then it will be important for him to keep in mind not only who he should try to appear to be in that context, but also who he really *is* in that context. After all, he needs to know his current strengths and weaknesses as an actor in order to know how to play his character well. He also needs to know his current thoughts and feelings in order to know how he relates to his character. For instance, insofar as he has the same thoughts and feelings as his character, he can channel those thoughts and feelings during his performance. But insofar as he has *different* thoughts and

feelings than his character, he will have to conceal his actual thoughts and feelings and merely “act”. And in order for him to give a good performance, it will be crucial for him to be able to tell the difference: to know when he can play his character well by speaking and acting authentically, and when he cannot.

Relatedly, if we have different personas and personalities, then it will help to keep in mind not only our *current* persona and personality, but also our personas and personalities in other contexts. After all, in order for a team of actors to put on a good collaborative performance, each needs to keep in mind not only his own character and strengths and weaknesses as an actor, but also his teammates' characters and strengths and weaknesses as actors. For instance, you may be able to do German accents, but if you initiate a scene where your entire team plays characters with German accents, you had better be sure that your teammates can do German accents too. Similarly, your "family personality" may be an especially good actor (because you feel especially comfortable around your family), and your family personality may also be especially disinclined to, say, have an affair. But if you want to play a game of poker, then the relevant question will not be, "How good of an actor am I right now?" but rather, "How good of an actor will I be *then*?" or in other words, "How good of an actor will the personality who comes out during the poker game be?" This will inform the "poker persona" that you decide to create for yourself, as well as how you play this character at the time. Similarly, if you want to, say, take a trip to Vegas, then the relevant question will not be, "How disinclined to

have an affair am I now?" but rather, "How disinclined to have an affair will I be then?" or rather, "How disinclined to have an affair will the personality who comes out during the trip be?" This will inform the "vacation persona" that you decide to create for yourself, as well as how you decide to play that character at the time (and also, of course, whether you decide to go to Vegas at all). Thus, if you have different personas and personalities, it will be important for you to have a conception of all of these personas and personalities at all times, so that you can initiate collaborative performances that work well for the whole "team."

Finally, in addition to having a conception for each of your characters and personalities, it will also be helpful for you to have a simple, idealized conception of who you are as a *person*, i.e., of who you are and what you care about across contexts. Why is this helpful? Because you will rarely be able to make decisions that just affect your current personality. Rather, the vast majority of the decisions that you make will affect the entire "team". Thus, it will be important for you to have a conception of who the team is and what the team wants, so that you can make decisions that represent a fair compromise among your teammates.

To see how this works, consider a typical family. Given how often our decisions affect our family members, it helps for us to have a simple, idealized conception of who our "family" is and what our "family" wants, so that we can make decisions that represent a fair compromise among our family members. For example, suppose that you prefer Chinese food, your spouse prefers Indian food, your kids

prefer pizza, and so on. As it happens, each of you hates the kind of food that the others prefer. But fortunately, you all like (not love, but like) Thai food. Thus, you decide that even though none of you prefers Thai food as a person, you nonetheless prefer Thai food as a family. In this kind of way, you can construct a self-conception of your family gradually, over time, by finding compromises among your family members, and then attributing these compromises as preferences to your family as a whole. Then you can identify with your family as a whole, rather than with yourself as an individual, when making decisions on behalf of your family members. For instance, if someone invites you all out to dinner and says, "What do you like?", the best response for you to give will not be "I prefer Chinese," or even "I prefer Chinese, my spouse prefers Indian...", but rather, "We prefer Thai." This process -- constructing a family self-conception that represents a fair compromise among your family members, and then identifying with this self-conception when making decisions on behalf of your family -- will make it much easier for you to make plans that your family members can also endorse, and help you to carry out.

Similarly: Given how often our decisions affect our other personalities, it helps for us to have a simple, idealized conception of who we are and what we want as a "person", so that we can make decisions that represent a fair compromise among our personalities. For example, suppose that your friend personality prefers staying out late, and your family personality prefers coming home early. Now suppose that a fair compromise between these preferences is: staying out for a while, and then

coming home in time to tuck your kids into bed. Thus, you might decide that even though none of your personalities prefers this plan, you nonetheless prefer this plan in general, i.e., as a person. In this kind of way, you can construct a self-conception of yourself as a person gradually, over time, by finding compromises among your personalities, and then attributing these compromises as preferences to yourself as a person. Then you can identify with yourself as a person, rather than with yourself as a personality, when making decisions on behalf of your other personalities. For instance, if someone invites you out for a drink and says, "What time do you want to get home by?", the best response for you to give will not be "I want to stay out late," or even "I want to stay out late now, but will want to get home early later...", but rather, "I want to get home by about nine." This process -- constructing a person self-conception that represents a fair compromise among your personalities, and then identifying with this self-conception when making decisions on behalf of your other personalities -- will make it much easier for you to make plans that your other personalities can also endorse, and help you to carry out. (Of course, if you, as a person, are also making decisions on behalf of your family, then you might have to repeat this process for this other, larger collective.)

This might seem like a very complicated process; but, as our participation in families, philosophy departments, and other collectives demonstrates, we do it quite naturally all the time. Moreover, note that, as I mentioned above, at least some philosophers and psychologists think that we already construct our person self-

conceptions in a way that allows them to play this role. Specifically, if we construct our person self-conceptions by finding our center of narrative gravity across contexts, then they will represent an "average" of our thoughts and feelings across contexts. For instance, if you construct your person self-conception by finding your center of narrative gravity, then you may well already think of yourself as a person who, "deep down," prefers to stay out for a while, but then get back in time to tuck his kids into bed. Of course, there is no guarantee that the "average" of your thoughts and feelings across contexts will amount to a fair compromise among them; and for that reason, it might be best for us to take control of our person self-conceptions by thinking more actively about what would be a fair compromise. Nevertheless, if Dennett is right about how we construct our person self-conceptions, then we can at the very least use the ones that we have as a good jumping-off point.

Granted, if we have multiple self-conceptions, then we might encounter some of the problems we considered above: we might conflate them together, and therefore become confused about who we really are, as well as what we should do and how we should do it. But keep in mind that most of these costs come not from *having* different self-conceptions, but rather from *conflating them together*. And in my view, our tendency to conflate them together comes primarily from our felt need to think of ourselves as more coherent, stable, and authentic than we really are – as well as from our unfortunate practice of using the same names and indexicals to refer to all these self-conceptions, in thought and in practice. This makes us likely to think of

our self-conceptions as competing answers to the same question: who am I, and what am I like?, rather than as different answers to different questions, e.g., what are my characters like? what are the actors who play them like? and what is this team of actors like in general?

If this is right, then we might be able to minimize the costs of having different self-conceptions if we switch to thinking about ourselves as a team of actors in a collaborative performance, and if we reinforce this in the way that we use language – e.g., if we use names like ‘my family persona,’ ‘my work persona,’ and so on to refer to ourselves as *characters*, ‘my family self,’ ‘my work self,’ and so on to refer to ourselves as *personalities*, and the first-person singular to refer to ourselves as *people* (though we will probably not be able to avoid using the first-personal singular to refer to ourselves as characters in interactions with others). This will make it easier for us to tell them apart, and to identify with the right self-conception in the right situation. i.e., to identify with our person self-conceptions when answering the question: What should I do with my day, week, life, etc.?.; to identify with our character self-conceptions when answering the question: How should I act right now, in order to play my role in this shared way of life?; and to identify with our personality self-conception when answering the question: how can I get myself to act this way? As a result, we will be less likely to be confused about what to do or how to do it, or to change our minds after making a decision. We will also be less likely to have experiences that resemble weakness of will, since we will know that some situations

require us to do what we think, at the time of action, is best under *some* self-conceptions but not under *others*.

Will thinking about our self-conceptions in this way eliminate all the costs of psychological fragmentation? Probably not. We will still encounter situations that make it hard to play all our roles well, as well as inner conflict that makes it hard for us to keep our commitments, or at least keep them wholeheartedly. Nevertheless, thinking about our self-conceptions this way can help us to at least minimize the costs of psychological fragmentation, thus making the cost-benefit analysis work out even more clearly in favor of (some degree of) incoherence and instability.

In sum: just as there are many benefits to living in a diverse society, there are many benefits to having a diverse psychology. If we have different personas, have different personalities, and have different self-conceptions, then we will be better off, individually as personalities (for the most part) as well as collectively as people. Granted, we will not have a fully coherent and stable set of beliefs, desires, aims, habits, and so on, and this will make our practical lives much harder, in many of the ways we have considered. But for many of us, at least, the benefits of this incoherence and instability will outweigh the costs. Not so much that we should be as fragmented as possible: just as we benefit from having different types of people in society, we also benefit from having many shared beliefs, values, practices, and traditions, so that we can keep the peace and let our differences make us stronger rather than destroy

us. Similarly, just as we benefit from having some incoherence and instability, we also benefit from having coherence and stability, so that we can keep the peace with ourselves and let our incoherence and instability make us stronger rather than destroy us.

Thus, in the intrapersonal case just as in the interpersonal case, we should neither try to maximize uniformity nor try to maximize diversity. States with violent sectarian conflict are still worse off because of their diversity, as are people with severe personality disorders. Instead, we should try to strike a healthy balance between integration and fragmentation, where this will likely involve more integration than fragmentation – though exactly how much integration and fragmentation is ideal will differ from state to state and person to person.

That said, we can probably safely say that, generally speaking, people will have to be integrated to a higher degree than families, states, and other groups of people. Not because people are any different in kind from these other groups, but rather because they are more *intimate* groups, and, as a result, their members will have to compromise and coordinate more often, and with respect to more issues. In other words, they will have to actively *get along* more, rather than just passively live and let live. Thus, they might benefit from having more shared beliefs, desires, aims, and so on than members of other, less intimate groups. Similarly, we can probably safely say that, given that a family is less intimate than a person but more intimate than a state, the members of a family might benefit from having more diversity than

a person, but also more unity than a state. And so on. Thus, we can see people as occupying one extreme on a spectrum of groups, where all groups benefit from a balance between unity and diversity, and the more intimate a group is, the closer to unity this balance will be. Though, again, this is only a very general prediction: there will probably be many exceptions in everyday life.

1.4. Practical Integrity

I think these considerations shed light on what we mean, or rather what we should mean, when we say that we should “integrate” or “unify” ourselves as rational agents. Specifically, we should understand this claim primarily in a *practical* sense, and only secondarily in a *psychological* sense. On this understanding, our primary aim, for each group that we are part of (whether this group is a person, family, state, and so on), should be to find a shared way of life that allows us to promote our individual and collective aims as much as possible. And then our secondary aim should be to try to share beliefs, values, practices, traditions, etc. with the other members of our group insofar as, and only insofar as, it helps us to do this.

I think that the interpersonal analogy is once again helpful here. We often say that a family, state, nation, and so on needs to be “united”. What do we mean by this? Not necessarily that every member of this group needs to be the exact same kind of

person, with the exact same beliefs and values. Rather, what we mean is that the members of this group need to be able to *act* as one – and to do it in a way that promotes our individual and collective aims as much as possible – whether or not they also *think* and *feel* as one. For larger groups, like states, this means democratically (or otherwise) electing "representatives" to make laws on behalf of the group, then following whatever laws they make. For small groups, like families, this means having the parents to make decisions on behalf of the family, and then the rest of the family acting on whatever decisions they make. In any event, however they compromise and coordinate, the upshot is that they will be able to act as one, and therefore they will also be able to *interact*, as one, with others. And this is, I think, what ultimately matters to us. Given how complex life is in a world with eight billion people, we need to organize into groups, and then interact as groups, in order to get by.³¹ Thus, for example, if a friend tells you that his family will be at your wedding, you need to be able to trust that they will be, even if some of them prefer not to. Similarly, if a state commits a crime against another state (or against its own people), we need to be able to trust that future representatives will take responsibility for this crime – even if some of them disagree with whether to have committed it.

As long as we can act, and interact, as one in this way, we are happy for the members of our group to be as psychologically diverse as they want to be. Indeed, when it comes to psychological unity/diversity, the only real question is: how much

³¹ Indeed, Goffman even claims that, in many social interactions, we should take the team as the *basic*, rather than the individual members of that team. (p. 80)

psychological unity/diversity does *practical* unity require, i.e., how much psychological unity/diversity will help us *act* as one in a way that promotes our individual and collective aims as much as possible? And, for all the reasons I discussed above, we tend to think that a balance between psychological unity and diversity will help us do this better than as much unity as possible. Thus, we think that, for many groups, practical unity requires at least a degree of psychological diversity: it requires having enough in common that we can get along (and live and let live), while at the same time being different enough that we can try out different forms of life, adopt a division of labor, compartmentalize harm and adapt to change, and benefit from an open exchange of ideas.

If my arguments in this paper are right, then we should say the same thing about people. When we say that we should “integrate” or “unite” ourselves as rational agents, we should take this to mean that we should be able *act* as one, in a way that promotes our aims as much as possible, whether or not we also *think* and *feel* as one. This may require a different strategy than in the interpersonal case: rather than task particular personalities with the role of speaking and acting on behalf of the person, we may have to make sure that *all* (or at least the vast majority)³² of our personalities are able to do this. In any event, however we do it, the upshot will be that we can act as one, and also that we can *interact*, as one, with other people. And in my view, this is

³² As Bloom and Carter both point out, some of personalities might be more like “adults” and others might be more like “children.” If so, then we might want to make sure that, as with a family, the “adult” personalities make the decisions and the “children” personalities just carry them out.

ultimately what we should care about at the intrapersonal level as well as at the interpersonal level. After all, given how complex life is in a world where *billions and billions* of personalities speak and act on behalf of the bodies they share, we need to organize into *people*, and then interact as people, in order to get by. Thus, for example, you need to trust, when a friend tells you that he will be at your party, that he really will be – even if his future self prefers not to go. Similarly, we need to trust, when a person commits a crime against someone else (or against himself), that he will take responsibility for this crime – even if his future self disagrees with whether to have committed it.

In my view, as long as people can act, and interact, as one in this way, we should be happy for them to be as psychologically incoherent and unstable as they want to be. Indeed, when it comes to psychological coherence and stability, the only question should be: how much coherence and stability does practical unity require, i.e., how much coherence and stability do we need to be able to *act* as one in a way that promotes our aims as much as possible? And, for all the reasons I discussed above, a balance between coherence and stability, on one hand, and incoherence and instability, on the other, will ultimately help us to do this better than as much coherence and stability as possible. Thus, for many people, practical unity requires at least a degree of psychological incoherence and instability (maybe not as much as requires does for other, larger groups; but still, it requires *some*): we should try to have enough coherence and stability that we can get along (and live and let live) with

ourselves, while at the same time having enough incoherence and instability that we can try out different forms of life, adopt a personal division of labor, compartmentalize harm and adapt to change, and benefit from an open exchange of ideas.

2. Is Self-Binding Morally Wrong?

“Over himself, over his own body and mind, the individual is sovereign.”

- John Stuart Mill

Imagine that you and your friend Doug are on a road trip, and you really want to get an early start tomorrow. (Doug is indifferent about whether to do this.) But you know that getting an early start tomorrow will be impossible if either of you drinks tonight, since any time either of you drinks, it takes both of you forever to get out the door the next day. Unfortunately, you also know that when night rolls around, both of you will want to drink. What should you do?

There are really two questions here. First what should you do about Doug? Second, what should you do about yourself?

We typically think that we need to ask and answer these questions in different ways, because morality places more constraints on your treatment of other people than it places on your treatment of yourself. Roughly speaking, we think that you have a duty to let other people live the way they want, whereas you have no such duty to yourself. In this case, for example, your options with respect to Doug are

limited. You can try to persuade him not to drink, and you can ask him not to drink for your sake. You can even offer to do something for him in exchange, like buy him a few extra drinks down the line. But you would be crossing a line if you did much more than this, for example if you took all his beer and money and kept it overnight, or if you locked him in your hotel room overnight, or even if you just threatened to do one of these things. We think that the former strategies for interacting with Doug – persuasion and compromise – are permissible because they respect his right to live the way he wants, whereas the latter – coercion and physical restraint – are impermissible because they infringe upon this right.

Of course, we can imagine cases where you might be permitted to coerce or physically restrain Doug all things considered, for example if you had reason to believe that he would drink and drive and thereby risk serious harm to other people. But we have to remember that these cases are the exception to the rule. You have to have a very good reason to justify coercing or physically restraining another person, and we can all agree that your desire to get an early start tomorrow morning is not good enough. If Doug chooses to drink tonight, despite your pleas to the contrary, then he may be a jerk, and a poor traveling companion. But he has every right to do this, and it would be wrong of you to stop him by force.

In contrast, we typically think that your options with respect to yourself are more expansive. For example, if you want to try to persuade yourself not to drink, or if you want to offer yourself something in exchange for not drinking, then you can do

that. But if you instead prefer to coerce or physically restrain yourself, by engaging in what philosophers call “self-binding,” then you can do that too.³³ For example, you can give all your beer and money to a friend to keep until morning, and then, if you really want to cover your bases, you can lock yourself in your hotel room overnight, or otherwise make it impossible for you to get to a bar. Far from infringing on your right to live the way you want, we typically think, this kind of behavior *permits* you to live the way you want, because it allows you to cultivate the kind of self-discipline and self-control that you need in order to take the means to your ends and preserve your integrity as a rational agent. We therefore conclude that self-binding is, if anything, to be encouraged rather than condemned.

I think that this conclusion is wrong. In my view, you have a duty to let *everyone* live the way they want, including yourself. It follows that self-binding is *prima facie* morally wrong: you need to have a very good reason to justify binding yourself, just as you do with other people. However, I also think that you can justify binding yourself more often, and with respect to more issues, than other people.

In this paper, I argue for these conclusions and explore some of their implications. First, I present a fusion case where you and Doug gradually turn from

³³ As Jon Elster describes, you can bind yourself “by removing certain options from the feasible set, by making them more costly or available only with a delay, and by insulating [yourself] from knowledge about their existence” (Elster, 2000, p. 1). An extreme example is when Ulysses tied himself to the mast, but we all do less extreme versions of this everyday. For example, we throw away food to keep ourselves from eating it. We put the alarm clock on the other side of the room to keep ourselves from hitting the snooze button. Some people even pledge to donate money to a cause they oppose if they get caught, say, drinking or gambling.

different people into psychologically continuous personalities in the same person – the “day self” and “night self,” respectively – and I argue that you still have a duty to let each other live the way you want even after this process is complete.

Second, I argue that this argument has surprising, but ultimately plausible, implications for real-world cases of self-binding.

Finally, I close by briefly considering the implications of this argument for justice and sovereignty. Specifically, I suggest that a person with multiple personalities has a just soul insofar as, and only insofar as, their personalities jointly commit to a fair set of laws; and a person has sovereignty, i.e., the right to live the way they want, insofar as, and only insofar as, they have a just soul.

2.1. Fusion

I begin by presenting an extended thought experiment in which you and Doug gradually turn from different people into different but psychologically continuous personalities in the same person.³⁴ The point will be to chart the course of your

³⁴ A note about my use of ‘psychological continuity,’ ‘personality’ and ‘person.’ First, I understand ‘psychological continuity’ in the same way that Parfit does. He defines ‘psychological connectedness’ as “the holding of particular direct psychological connections” and ‘psychological continuity’ as “the holding of overlapping chains of strong connectedness.” (1984, p. 206) On this usage, psychological connectedness comes in degrees: A is psychologically connected to B *to the degree that* A shares particular direct psychological connections with B. In contrast, psychological

moral relationship throughout this process. When, if ever, do you lose your duty to let Doug live the way that he wants, and when, if ever, does he lose his duty to let you do the same? And my answer will be: this never happens. However, I will also argue that as your fusion proceeds, you and Doug become permitted to interfere with each other more all things considered. Thus, this thought experiment will really be a story about how you have the same kind of moral relationship with your “night self”

continuity is all or nothing: A is psychologically continuous with B *if and only if* A shares an overlapping chain of strong psychological connectedness with B. (Parfit defines psychological continuity this way in order to make it a suitable criterion for personal identity, which is also, he thinks, all-or-nothing.)

Parfit argues that psychological connectedness and/or continuity with any cause is “what matters in morality,” or in other words, what gives A characteristically self-regarding duties to B instead of characteristically other-regarding duties to B. Part of my goal in this paper, however, is to argue that A can have characteristically other-regarding duties to B even if they share a *very high degree* of psychological connectedness and continuity.

Second, some psychologists use ‘personality’ to refer to all of a person’s beliefs, desires, aims, character traits, and so on. But others use it in a more restrictive way, to refer to a *coherent and stable set* of beliefs, desires, aims, character traits, and so on. (Carter 2008, p. 18-9) On this latter usage, a person can have multiple, psychologically continuous personalities. For example, I can have one personality in my twenties and another in my thirties, or I can have one personality at work and another at home. I will use ‘personality’ in this latter way.

Third, some philosophers use ‘person’ in a normative sense, or as they sometimes put it, in a “forensic” sense, to refer to the “basic unit of moral concern.” (For example, see Korsgaard 2009, Rovane 1998, and Schectman 1996.) Others, in contrast, use the term in a descriptive sense, to refer to a certain kind of thing *whether or not* it counts as the basic unit of moral concern. (For example, see Parfit 1984.) I will use ‘person’ in this latter, descriptive sense. Moreover, I will assume that if A has the same body, the same brain, and a very high degree of psychological connectedness and continuity with B, with no competitors, then A is the same person as B. (This is an uncontroversial assumption, even, it seems, among people who use ‘person’ in the former, normative sense.)

that you have with other people, even though your duties to your “night self” are still, in practice, very different from your duties to other people.

Before I begin, let me make an important clarification. Since my strategy is to take you and Doug through a series of changes, and argue that none of these changes is sufficient to take away your duties to each other and rights against each other, my argument might appear to trade essentially on a sorites case. But this appearance would be misleading. The idea is not to take you through a series of small changes and ask, each step of the way, whether a change *this small* could possibly take away your duties to each other and rights against each other. (Indeed, each of these changes will be very big.) Rather, the idea is to isolate each of the main differences between your relationship with other people, on one hand, and your relationship with your “night self”, on the other hand, and to argue that none of these differences is morally relevant for present purposes. You have a duty to let Doug live the way he wants, and vice versa, not because you have different bodies or even different streams of consciousness, but rather because you each have ends to pursue, and sometimes your ends may conflict; and this remains true even when you are parts of the same person. Indeed, I will argue that if anything, sharing a body with Doug makes your duties to him and rights against him all the more *important*, and sharing psychological connectedness and continuity with him has the same effect, although it also, fortunately, makes your duties to him easier to discharge, and your rights against him easier to protect.

The Set Up

Imagine that you are a workaholic. You like to wake up early and make a light breakfast while you listen to NPR. Then you like to get into the office, so you can get some work done before the day starts to get away from you. You love your job, so you take on a lot of extra commitments, and you fill up your days with emails, meetings, coffee breaks, and other forms of professional activity. This typically wipes you out, so at the end of the day you like to go home and have a quiet night by yourself; you cook dinner and watch TV or read a book. The one exception is: every now and then you invite some friends from work over for dinner. You eat good food, drink good wine, and maybe watch a movie or play a card game. Otherwise, though, you rarely drink, and you almost always get to bed early, because you hate feeling sluggish and hung over in the morning; you like to be able to jump out of bed and do it all again.

Now imagine that Doug is different from you in all of these respects. He likes to wake up late and eat ice cream for lunch. Then he spends a few hours watching TV, usually with a pack of cigarettes and a six-pack of beer. After that, he might do any number of things. He works as a bartender and plays in a punk rock band; so he might go to work, or to band practice, or to the recording studio. Then, after his work is done, Doug often goes out on his own: he goes to bars and clubs and shows,

and then he wanders around the city and talks to strangers until four or five in the morning. Finally, on occasion, he likes to invite a bunch of people over to his apartment for a party – not a dinner party, but a real party. He puts on some music, and they all drink and dance until they pass out, or go to the diner for some breakfast.

You like Doug a lot, even though you fundamentally disagree with him about how to live. In fact, you like him so much that at one point you thought that you had more in common than you actually do. That is, you thought that deep down, Doug must really want the kind of life that you have. But the more you got to know Doug, the more you realized that, no, he really does want the kind of life that *he* has. As a result, you would never dream of doing anything to interfere with his life choices. You might not identify with them, but you also recognize that Doug has a right to live the way he wants, and that to somehow force him to start waking up early and going into an office would be morally wrong: it would be paternalistic at best and tyrannical at worst.³⁵ And of course, you also think that you have the same rights against Doug that he has against you, and that even if he sincerely believed that you

³⁵ I should distinguish two kinds of paternalism here. A weak paternalist says that you can permissibly interfere with someone for his own sake iff he takes the wrong means to his ends. A strong paternalist, in contrast, says that you can *also* permissibly interfere with someone for his own sake if he chooses the wrong ends. (See Dworkin 2005.) This is relevant because I am stipulating that you and Doug know what your ends are, as well as what the means to those ends are. Therefore, you can permissibly interfere with Doug for his own sake, and vice versa, only if strong paternalism is true. And few philosophers think that strong paternalism is true.

should start staying out late drinking, he would be wrong to in any way *force* you to do that. And fortunately for you (and for your relationship with Doug), he agrees with you about this.

Step 1: Sharing an apartment

Now imagine that economic circumstances force you and Doug to move into an apartment together.³⁶ You keep such different hours that you rarely ever see each other. Nevertheless, you start to clash a lot more than you used to. When Doug has people over at four in the morning, you have a hard time sleeping. When he leaves a mess in the kitchen at the end of the night, you have a hard time making breakfast in the morning. When he decorates the apartment with disco balls, punk rock posters, and empty liquor bottles, you have a hard time presenting yourself as a serious professional to your dinner guests. And so on. Similarly: when you wake up early and turn on NPR, Doug has a hard time sleeping in and nursing his hangover. When you insist that he has to clean up after himself every night, he has a hard time having fun with his friends. When you decorate the apartment with potted plants, coffee-

³⁶ I mean this in a very strong sense of ‘force.’ Neither of you has any other option whatsoever. You either live together or live on the street and freeze to death. This is important because it makes this stage in your fusion relevantly similar to later stages, which also feature a lack of any meaningful choice on your part.

table books, and furniture sets from IKEA, he has a hard time presenting himself as a counter-cultural punk to his friends. And so on.

How does this affect your moral relationship with Doug?³⁷ Well, you still have a duty to let him live the way he wants, and a right to do the same. But now your relationship has changed in a way that makes your duties to each other harder to discharge, and your rights against each other harder to protect. Before you moved in together, it was easy for you to let Doug do what he wanted, and vice versa: all you had to do was get out of the way. You were free to spend as much time apart as you needed to, in order to live and let live. But now that you share an apartment, all of this has changed. You both spend a great deal of time in your apartment, and you also do many of the things that you most enjoy there. Moreover, your apartment is an important part of your self-presentation³⁷: the way that you present your living space to others sends a message to them about who you are, and this affects not only how they think about you, but also how you think about yourself.³⁸ Thus, since your apartment is essential to your ability to pursue your conception of the good life, and since you and Doug have such different conceptions of the good life, your living together makes it impossible for both of you to live *exactly* the way you want. So now

³⁷ See Goffman 1959, particularly pp. 106-141.

³⁸ As labeling theorists in sociology are quick to point out, the way that we think about ourselves depends on how others treat us, and the way that others treat us depends on how *they* think about us. So, for example, if a teacher thinks of a student as a troublemaker, then he might start *treating* his student that way; and if so, then his student might start thinking of *himself* as a troublemaker, and act the part. See, for example, Goffman 1959, Swann 1999, and Velleman 2009.

the stakes are raised: if either of you lives exactly the way that you want, the other will be incapable of doing the same.

Thus, you and Doug both have to work harder, and sacrifice more, in order to discharge your duties to each other now. What does this mean in practice? Given how intertwined your lives are, you have to talk through your problems, find a mutually acceptable set of “apartment policies,” and then jointly commit to following those policies. For example, you may prefer to buy a new furniture set from IKEA, and Doug may prefer to scrounge old furniture off the street. But maybe you can find a middle ground, e.g., you can agree to fill your apartment with decent used furniture. Of course, neither of you will *love* this option, but you might like it well enough, and you should both be able to fold it into your self-presentations.³⁹ Similarly, maybe you can let Doug decorate the apartment however he likes, provided that you can put some of his things away when you have company over, and maybe you can let him have friends over late at night, provided that they keep the noise down and clean up after themselves.

³⁹ Will a fair compromise give you and Doug equal opportunity to live the way that you want? Probably not. Doug’s ideal lifestyle threatens your interests more than your ideal lifestyle threatens his. For example, if you kept your apartment dirty and filled with cigarette smoke, as Doug wants to do, then this would pose a serious threat not only to your aesthetic sensibilities but also to your *health*, whereas if you kept your apartment clean and smoke-free, as you want to do, then this would, at worst, pose a serious threat to Doug’s aesthetic sensibilities. Thus, a fair compromise will probably require Doug to sacrifice more than you – not because your ideal lifestyle is in any objective sense “better” than his, but rather just because, unfortunately for Doug, his ideal lifestyle would cause more collateral damage than yours, and therefore he needs to work harder than you in order to discharge his duties to you.

Ideally, then, you and Doug would jointly commit to a fair compromise, and both of you would live up to your commitments. But realistically, you can expect each other to fall short now and then; and if this happens, then intuitively, you have a right to interfere with each other in order to enforce your commitments. For example, if Doug plays loud music every night, then you have a right to ask him to keep it down, and in extreme cases (e.g. if you have a very important meeting in the morning) to *make* him keep it down, for example by taking away his stereo for the night. And if you play the radio loudly in the morning, then Doug has a right to ask you to keep it down, and in extreme cases (e.g. if he has a very important show that night) to *make* you keep it down, for example by taking away your radio for the morning. This is of course a recent development: before, it would have been intrusive (and weird) for you to interfere with Doug in these ways, and vice versa. Why should you care if Doug listens to music at night? And why should he care if you listen to the radio in the morning? Live and let live. But now that you live *together*, you can interfere with each other in these ways without being intrusive – not because you can do whatever you want to each other now, but rather because you may *have* to interfere with each other in these ways in order to protect your own rights.

Now, does any of this call into question the fact that you still have a duty to let Doug live the way that he wants, and vice versa? Of course not. On the contrary, this entire discussion has been about how you can continue to do this, in light of your

new, more intimate living situation. Nevertheless, your moral relationship with Doug has changed in an important way: you now clash more than you used to, and therefore you have to compromise and coordinate more. And that means that, first, you have to make greater sacrifices for each other now, and second, insofar as one of you fails to do this, the other has a right to interfere with him in order to *make* him do it (but not in order to make him do anything more than that).

Step 2: Sharing a body

Next, imagine that an evil scientist kidnaps you and Doug, takes complete brain scans of both of you, and destroys your bodies. He then builds a new body and gives this body multiple personality disorder.⁴⁰ Specifically, the evil scientist makes it the case that from six in the morning to six at night, this new body has *your* personality, and from six at night to six in the morning, he has *Doug's*. This means that for half of each day, this new body is fully psychologically connected to and continuous with you: he has all of your beliefs, memories, anticipations, desires, aims, intentions, habits, and so on.⁴¹ And for the other half of each day, he is completely psychologically connected to and continuous with Doug, in all of the same ways. And

⁴⁰ Many psychologists now prefer the term 'dissociative identity disorder,' or 'DID' for short, to describe this condition. See Hacking 1995 for a good philosophical discussion of DID.

⁴¹ I am using 'psychological connectedness and continuity' in the same way as Parfit 1984. See note 3 for more on this.

that means that this new body's two personalities are completely psychologically *disconnected from and discontinuous with* each other: they have no beliefs, memories, anticipations, desires, aims, intentions, habits, and so on in common, so each "blacks out" when the other takes over. When the clock strikes six in the morning, Doug's personality goes to sleep and yours wakes up, and when the clock strikes six in the evening, the opposite happens.⁴²

How does this change your relationship with Doug?⁴³ Well, you live together in a much more intimate way than before. Your apartment may be where you spend most of your time, and where you present yourself to your guests; but ultimately you can leave your apartment and find other settings in which to present yourself to people if you need to. But your body is where you spend *all* of your time, and where you present yourself to *all* the people in your life.⁴⁴ Thus, the way that you treat your body affects nearly every aspect of your life, and that means that even though you and Doug never see each other at all now, you still clash a lot more than you used to. For example, if Doug stays out until six in the morning, you will have *no chance* of

⁴² In what follows, I will refer to your personality as 'you' and Doug's personality as 'Doug.' This is meant to be neutral on questions of personal identity.

⁴³ Obviously, a lot depends on what kind of new body you have. If your new body is very different from your old one, then this will have a profound effect on your experience of yourself and the world. But my interest is not in how having a new body would change you, but rather in how sharing a body with Doug would change your moral relationship with him. So if you like, you can imagine that you and Doug were identical twins before your fusion, and that your new body is qualitatively identical to your old one.

⁴⁴ One partial exception is your activity on the internet or telephone: you can interact with people in these ways while making only minimal presentational use of your body.

waking up early and having a productive day. If he gets tattoos and piercings all over his face, you could get fired from your job. If he smokes too many cigarettes, you could end up with cancer. And so on. Similarly: if you wake up early and have a productive day, then Doug will have *no chance* of staying out late drinking. (We can imagine that your work is very taxing.) If you drink too much coffee, he might not only crash later in the day, but also suffer from high blood pressure later in life. And so on. Before you and Doug could do these things and endanger only yourselves (give or take), but now you would be endangering each other too – and in a much more direct way than you could have done previously – with the result that your lives are much more intertwined than they used to be. So now, if either of you tries to have exactly the life that you want, then the other will have *nothing* like the life that he wants.

So what should you do? Practically speaking, you have two options. Your first option is an extension of what you did before: you can work together to find a new, more expansive set of “bodily policies” and then jointly commit to following them. For example, you might prefer a normal haircut and Doug might prefer a green Mohawk. But maybe you can find a middle ground, e.g., you can get your hair cut short enough for you to wear it normally at work but long enough for Doug to spike it at the bar. Similarly, maybe you can let Doug get some tattoos and piercings, provided that you can cover them up when you go to work, and maybe you can let

him stay out late drinking, provided that he gets home by two in the morning and has plenty of water before bed.

If you and Doug take this first option, then both of you will still have something resembling the life that you want. But it will be a difficult life, since, first, you will have to compromise and coordinate more often, and with respect to more issues, than before, with the upshot that very few aspects of your life will be up to you alone. And second, your negotiations will be a lot harder to carry out than they were before: instead of talking to each other face to face, as you might have done when you had different bodies, you will have to leave notes for each other, writing down your proposals and counter-proposals and then, ultimately, your agreements and promises, over the course of many days. But we can imagine that you and Doug are patient enough to make this work.

Your second option, however, is to *make* Doug live the way that you want. How can you do this? By making it impossible, or at any rate very difficult, for him to live the way that *he* wants, and then by threatening to harm him even more unless he does what you say. For example, suppose that you wake up one morning and find a new message from Doug, outlining a new set of ideas about how to get along. But instead of considering his ideas and writing a response, as you normally do, you decide to take a different approach: You pour all of his beer down the drain, tear all of his posters off the wall, and smash all of his records into a thousand pieces. Then you take all of his money out of the sock drawer and put it in your bank account, take

all of his clothes out of the closet and donate them to the Salvation Army, and take all of his recording equipment and sell it on e-Bay. Then you call his boss and say that you quit, and call his bandmates and say the same (since, after all, you have the same voice now). Finally, you send Doug an email in which you reject his peace plan and say this instead: if he even thinks about staying out late tonight, you will see to it that he never steps foot in a bar or club again.⁴⁵

The question is: What does morality have to say about this? Are you morally permitted to do this on the grounds that you share a body with Doug now? It seems clear that the answer is no. Instead, sharing a body seems to have changed your moral relationship in the same kind of way that sharing an apartment did: you still disagree about how to live, but now you live together in a much more intimate way than you did before. Thus you have to work much harder in order to discharge your duties to each other and protect your rights against each other. And this has two familiar upshots: first, you have to make greater sacrifices for each other now, and second, insofar as one of you fails to do this, the other has a right to interfere with him in order to *make* him do it (but not in order to make him do anything more than that).

⁴⁵ Of course, if you take this route, then Doug may decide to strike back, and then you may have to strike back, and so on, with the result that you eventually get stuck in a civil war that makes *both* of you miserable. And keep in mind that the costs of civil war are greater now that you share a body, since the winner, if there is a winner, will have to suffer any bodily damage that he inflicts on the loser, and the loser will have to endure his loss forever, rather than starting over someplace new.

Step 3: Sharing cognitive attitudes

Now imagine that the evil scientist returns. He kidnaps you and Doug and opens up a small psychological “tunnel” between you. This means that you and Doug are no longer completely psychologically disconnected. You can remember his past experiences and anticipate his future experiences, and vice versa.⁴⁶ And as a result, you and Doug become psychologically connected in other respects too. For example, if Doug forms a belief on the basis of some experience, then you are likely to inherit his belief as well as his memory of the experience. But imagine that your psychological connectedness stops there. This means a few things. First, even though you share many beliefs, memories, and anticipations with Doug, you also keep some to yourself, and so does Doug. Second, with respect to the beliefs, memories, and anticipations that you *do* share, you may rearrange them in your web of mental states, so that a belief, memory, or anticipation that is central for Doug may be quite peripheral for you.⁴⁷ Third, and most importantly, you still keep all of your *conative*

⁴⁶ Some philosophers prefer to use ‘quasi-memory’ to refer to a state which is neutral with respect to whether the remembering subject is the same person as the subject whose experience is being remembered. I believe that you and Doug become the same person by the end of this fusion, but I want to remain neutral on precisely when this happens. So if you prefer, you can read me as saying that you share quasi-memories, quasi-anticipations, and so on at this stage.

⁴⁷ Technically speaking, these qualifications are not necessary for my purposes here. My main argument would still work if you and Doug were fully cognitively connected. But I add these qualifications in order to make you and Doug more like actual personalities, which are usually mostly, but not fully, cognitively connected in

attitudes to yourself, and vice versa, i.e., you keep all of your desires, aims, plans, and so on, and Doug keeps all of his.⁴⁸

In many ways, this change is just as radical as the last one. Now you and Doug are linked by a chain of memory and anticipation, and this means that the black outs stop: you remember *being* Doug, and Doug remembers *being* you.⁴⁹ As a result, you can now use different methods for interacting with each other. Before, you had to use “third-personal” methods for interacting with Doug, like leaving notes for him in prominent places. But now you can use “first-personal” methods for interacting with him, like forming an “intention” to act, with the hope that he remembers the intention and acts on it. (I put ‘intention’ in scare quotes for a reason; more on this in a moment.)

Nevertheless, you and Doug still disagree about how to live. When six in the evening rolls around, Doug may remember everything that you said and did over the past twelve hours, as though *he* said and did it. But he also (seems to) experience a

precisely these ways. For example, my “work” personality remembers more of my work experiences than my “home” personality does, and he also remembers them more vividly. Psychologists refer to this phenomenon as “state-dependent recall.” (See, for example, Carter 2008, pp. 45-73.)

⁴⁸ Roughly speaking, a cognitive attitude has a “mind-to-world” direction of fit, i.e., the point of the attitude is to represent the world as it *is*. In contrast, a conative attitude has a “world-to-mind” direction of fit, i.e., the point of the attitude is to represent the world as it *ought to be*. So on this usage, attitudes like beliefs, memories, and anticipations are usually cognitive, and attitudes like desires, aims, and plans are usually conative. But for my purposes here, I will also count *beliefs about value* as conative attitudes. So at this stage in your fusion, you still keep all your beliefs about goodness, badness, rightness, wrongness, and so on to yourself, and so does Doug.

⁴⁹ Again, I mean to be neutral at this stage about whether you *are* Doug, in the metaphysical ‘personal identity’ sense.

sudden change in attitude: everything that (seemed to) matter to him only moments before no longer does. He remembers waking up early and listening to NPR, he remembers going into work and writing emails and having meetings, he remembers coming home and reading poetry, and he remembers *enjoying* all of these things. But he also feels thoroughly alienated from these remembered experiences: he remembers each activity “from the inside,” i.e. from the standpoint of the person who did it and enjoyed it, but he also has trouble making sense, intuitively, of why one would do it or enjoy it. Because from his standpoint now, it seems clear that these activities are completely trivial. What really matters, what really makes life worth living, is meeting new people and having new experiences. As a result, even though Doug remembers your experiences *as* pleasurable, he is unlikely to take much pleasure in his memory of them.⁵⁰ And even though he remembers your “intention” to stay in and go to bed early, he is unlikely to *act* on this intention unless you give him some other reason to – a reason that speaks to his current evaluative attitudes.

⁵⁰ This is an important point. There are two senses in which a memory (or anticipation) can be pleasurable (or painful). The first is: it can represent its *content* as pleasurable. And the second is: it can *itself* be pleasurable. In my view, these can come apart. First, you can remember an experience as pleasurable without taking any pleasure in the *memory*, for example, when a person thinks about his past infidelity with a sense of regret. And second, you can take pleasure in a memory without remembering the *experience* as pleasurable, for example, when a person laughs at his memory of being terrified of roller-coasters.

It may be that, in order to *fully* remember (or anticipate) an experience, you have to identify with the protagonist of this memory enough to feel his pleasure and pain (or at least not feel the opposite). If so, then we can think of your memories at this stage as *partial*, to the degree that they alienate you from the subject of your remembered experience.

Moreover, you feel the same alienation from *his* evaluative attitudes, and the same reluctance to act on *his* “intentions”, despite remembering them all, when you take over again the next day.

This means that even though you and Doug can now use first-personal *methods* for interacting with each other, you still have to use third-personal *strategies* for interacting with each other, i.e., you still have to give each other an *incentive* for acting on the “intentions” that you pass back and forth. (This is why I put ‘intention’ in scare quotes before. A real intention is supposed to motivate all by itself, without any extra incentives.)

Practically speaking, then, you and Doug still have the same two options that you had before. Either you try to get *some* of what you want by compromising with each other, or you try to get *all* of what you want by coercing each other. Fortunately, compromising with Doug will be much easier now. Your ability to use first-personal methods for interacting with each other will streamline your negotiations considerably. Think about everything that you had to do in order to compromise with Doug before. First you had to think about what you wanted, then you had to think about what he wanted, and then you had to think about what would make for a fair compromise between you. Then you had to think about your negotiation strategy. How should you frame the debate? How should you argue for your position? Then, once you settled on your strategy, you had to put everything into language, and write it all down. And of course, you also had to try to interpret

Doug as he did all the same things. What is he trying to say? Is he making a persuasive argument? Should you take him at his word? And so on. But now that you share memories and anticipations with each other, you can dispense with all of this. All that you have to do is think about what you want, what Doug wants, and what would make for a fair compromise between you; and then you can rely on Doug to inherit your thoughts and do the same. Or, if you really think that your idea is fair (and that Doug will agree), then perhaps you can skip your negotiations entirely and express your idea in the form of an “intention”⁵¹: for example, you can “intend” to have a couple of drinks tonight and then come home at two in the morning, with the expectation that Doug will remember your “intention” and act on it – not because he *wants* to do this (he prefers to drink more and stay out later), but rather because he recognizes that you formed this “intention” in the spirit of compromise.⁵²

⁵¹ I continue to put ‘intention’ in scare quotes because one might think that an intention is a conative state, and therefore, according to my stipulation, you and Doug are incapable of sharing intentions at this stage in your fusion. But I think that, if this is right, then whether or not you share an *actual* intention is irrelevant: all that matters is that you can decide what Doug will do. Whether he complies with your decision because (a) he inherits your intention or (b) he remembers your decision first-personally and regards it as authoritative, makes no practical (or, I suspect, phenomenological) difference.

⁵² Another, even more effective way of coordinating your activity with Doug is to form a shared self-conception, and then to form intentions from the standpoint of your shared self, rather than from the standpoint of your individual selves. We do this all the time in regular life. For example, a parent might decide what to make for dinner by thinking about what “we” want to eat rather than by thinking about what “I” want to eat. And I think that, once you and Doug start to share cognitive attitudes, it will be much easier, and also much more important, for you to do this often, if not all the time. See my “Practical Integrity” for more on this.

Moreover, the same psychological transparency that makes compromising with Doug easier will also make coercing him harder, for three main reasons. First, Doug knows all your plans now, and this means that he can anticipate your attacks (provided that you start planning at least a day in advance) as well as tell whether your threats and promises are credible. Will you really throw away all of his records if he throws away all of your books? Or conversely: Will you really stop attacking him if he stops attacking you? Second, Doug has more leverage against you now. Because in addition to being able to steal your money, destroy your possessions, and so on, he can also plunder your psychology for your deepest, darkest secrets and exploit them. Third, you may have a hard time hurting Doug now without also, indirectly, hurting yourself, since you may have a hard time enjoying the thrill of victory if you also have to remember and anticipate the agony of defeat everyday – especially if you have any sense of empathy at all. Granted, all of these changes to your relationship will also make it harder for Doug to coerce you. But still, the upshot is that coercion will be harder now; and if either of you attacks the other, then you might very quickly find yourselves stuck in a civil war that destroys both of your lives *completely*.

All things considered, then, you and Doug now have even more reason to compromise rather than coerce. But suppose that one of you decides to try to coerce anyway. For example, suppose that one day, Doug decides that he can beat you in a civil war as long as he strikes now, before you wake up and remember his plans. He

knows that this will make you miserable. But he thinks that he can remember and anticipate your misery without feeling miserable *himself*, since your misery will be based on values that he utterly rejects. Thus, instead of acting on the “intentions” that you formed in the spirit of compromise, he goes on the offensive. He cuts up all your clothes, shaves your head into a Mohawk, and gets tattoos and piercings all over your body. Then he takes all of the money out of your bank account and spends it on expensive (and non-returnable) recording equipment. Then he sends your boss an email saying that you hate him, and you quit. Finally, he stays up till six in the morning, drinking and doing a cocktail of drugs that will make you catatonic the next day, if you even wake up at all. And he continues on like this, day in and day out, until your ability and inclination to strike back fades away entirely, and you relent, miserable but out of options.

Once again, the question is: What does morality have to say about this? Is Doug morally permitted to do this to you, on the grounds that you share beliefs, memories and anticipations with him now?⁵³ Intuitively, the answer is no. In fact, if

⁵³ A related question is: do you and Doug have a duty to “unify” yourselves, by conditioning yourselves to have the same beliefs, desires, aims, and so on, across time? I discuss this question in detail in “Practical Integrity,” and I also address it briefly below. But roughly speaking: my view is that you do not have a duty to unify with Doug in this sense. In fact, you may have good reason to stay psychologically *fragmented*, since this allows you to preserve your own identity, see the world from different perspectives, experiment with different ways of living, and adopt a personal division of labor, where each personality performs different tasks. (Just to be clear, I mean ‘identity’ here in the characterization sense, not in the reidentification sense. See Schectman 1996 for discussion about this.) However, as I argue below, I believe that you *do* have a duty to unify with Doug in a different sense: you have a duty to

anything, sharing cognitive attitudes seems to have changed your moral relationship in the same kind of way that sharing an apartment and a body did, because now you share *psychological* space in addition to *physical* space, and therefore your lives are even more intertwined than they were before. And that means that, first, you have to make greater sacrifices for each other than you did before, and second, you have a right to interfere with each other more than you did before, in order to protect your rights against each other. For example, if it bothers you that Doug can remember everything that you say and do, then you have a right to ask him to be discreet with this information. And if it bothers you that you can remember everything that *Doug* says and does, then perhaps you have a right to ask him not to say and do the things that prove especially distracting and troubling for you. Fortunately, as I said before, these negotiations will be much easier to carry out, now that you can talk to Doug without ever opening your mouth, or opening your email, or picking up a pen. But they will also have to cover pretty much every aspect of your lives now, with the result that morally speaking, pretty much *nothing* is up to you alone anymore.

Step 4: Sharing conative attitudes

ratify a “self-constitution” or, in other words, to jointly commit to a fair set of laws, either by compromising with each other or by forming your intentions from a shared practical perspective.

Imagine, finally, that the evil scientist kidnaps you and Doug again and makes you even more psychologically connected. Now, in addition to sharing certain beliefs, memories, and anticipations, you also share certain desires, aims, and plans. But imagine that this sharing is still imperfect, and in the same kind of way as before: you still keep some conative attitudes to yourself, and you “rearrange” the conative attitudes that you share in your web of mental states. So for example, a certain aim may be central for Doug but peripheral for you, and vice versa.

How does this change your relationship with Doug? You now share a very high degree of psychological connectedness, and that means you not only remember and anticipate his experiences but also *identify* with some of them more. For example, you wake up in the morning and remember what Doug did last night, as though you did it: you remember going to the bar, having some drinks, playing a show, and then wandering around the city until two in the morning. And you also remember *enjoying* all of these things. But whereas before you took no pleasure in your memory of these experiences, now you take at least some. For example, you now fully endorse being in a punk rock band. You like the music, you like the people, and you like the (remembered) thrill of performing live. However, you still prefer not to drink much. You know that it makes Doug happy, but you have no interest in doing it, nor do you take any pleasure in your memory of doing it (despite your remembering it *as* pleasurable). And in the same kind of way, Doug identifies with some of your experiences more than he did before, while still feeling alienated from others.

In some respects, this will change your relationship with Doug significantly. Because with respect to the issues that you agree about, you can rationally use first-personal methods *and* strategies for interacting with each other, which will once again streamline your interactions considerably. For example, suppose that you and Doug both want to go on vacation. And suppose that you both agree on all the general parameters: where to go, when to go, how much money to spend, and so on. How should you proceed with respect to this issue? Easy. Just start planning your trip. For instance, you can form an intention to research hotels online tonight, with the expectation that Doug will remember your intention and act on it. Then Doug can research hotels, pick a room, and make a reservation without consulting you first, because he knows that with respect to this issue, *his* decision is *your* decision. Then Doug can form an intention to rent a car tomorrow morning with the same expectation, and you can act on this intention without consulting him first, for the same reason. And so on. In short: when it comes to planning your trip, you and Doug can act as one. You can set ends for yourselves and then take the means to them, and you can do this without having to compromise at all. All you have to do is coordinate your behavior over time, by forming and acting on intentions.⁵⁴

⁵⁴ Luca Ferrero 2010 argues that making and acting on long-term intentions is rational because it allows your temporal selves to adopt a division of deliberative labor. I think this is exactly right. In this case, both you and Doug will be better off if you trust him to decide which hotel to book and if he trusts you to decide which car to rent. Granted, you *might* have made a different decision than Doug if you looked at all the information again, yourself, but you have no real reason to believe that you would have (or that it matters much one way or the other); and you and Doug both

In other respects, however, your relationship with Doug is exactly the same as it was before. Because with respect to the issues that you still disagree about, you still have to use third-personal strategies for interacting with each other. For example, even though you and Doug agree that you should go on a trip, and even though you agree about the general parameters of the trip, you also disagree about many of the specifics. For example, you really want to get an early start on the first day of the trip. (Doug is indifferent about this.) However, you know that getting an early start will be impossible if Doug drinks the night before, since anytime Doug drinks, it takes you forever to get out the door the next day. Unfortunately, you also know that when the night rolls around, Doug will want to drink. So what should you do?

Well, since you and Doug still disagree about whether to drink, you still have the same two options as before. First, you can try to compromise. For example, you can ask Doug not to drink tonight for your sake, and you can promise to let him have a few extra drinks tomorrow night in exchange. And, as always, if you really think that your offer is fair (and that Doug will agree), then you can skip your negotiations entirely and express your offer in the form of an intention: you can intend not to drink at all tonight, and then to drink *a lot* tomorrow night, with the expectation that Doug will remember your intention and act on it. Alternatively, if you want to get an early start tomorrow morning without having to give up *anything* in exchange, then

have a strong interest in deferring to each other in this kind of way, given how many decisions you have to make and act on together everyday.

you can accomplish this by attacking and threatening Doug. For example, you can take all his beer and pour it down the drain, take all his money and give it to a friend to keep until morning, and then lock yourself in a room until your friend returns. Then you can tell Doug, merely by framing the intention in your own mind, that if he so much as *thinks* about drinking tonight, you will do everything in your power to make sure that he stays sober the whole rest of the trip, and maybe for much longer than that.

The question is, as always: What does morality have to say about this? Are you morally permitted to do this to Doug on the grounds that you now agree about what to do *with respect to other issues*? Intuitively, the answer is still no. The fact that you now agree with Doug about, e.g., what kind of hotel to stay at, or what kind of car to rent, in no way frees you to do whatever you want to him otherwise. Granted, you are now morally permitted to book the hotel of your choice, or rent the car of your choice. But this is only because you and Doug have the same preference with respect to these issues, and therefore you can do what you want without interfering with each other at all. However, if you were to change your mind about these issues, then you would once again have to compromise with Doug. And you also, of course, have to compromise with him with respect to the issues that you never agreed about in the first place. In short, then: you and Doug still have a duty to let each other live the way that you want. Fortunately, insofar as you agree about how to live, you can do this without sacrificing your own interests at all. But insofar as you still disagree

about how to live, you still have a duty to make sacrifices for each other, as well as a right to bind each other only insofar as you have to, in order to protect your rights against each other.

Step 5. Sharing a self-conception

Now imagine that the evil scientist comes back one final time, to bring about one final change in your relationship with Doug. Specifically, he makes it the case that you and Doug share a self-conception. Whereas before you used the first-person plural to refer to you and Doug, as a pair, you now use the first-person singular to do this. And whereas before you used the first-person singular to refer to yourself, as a personality, and the second- or third-person singular to refer to Doug, as a personality, you now use "my day self" to refer to yourself and "my night self" to refer to Doug.

In a sense, this is just a minor change. You have the same exact relationship with Doug that you did before. Moreover, you have the same exact self-conceptions that you had before. Before the change took place, you had *individual* self-conceptions, i.e., conceptions of who you were, and what you were like, as personalities. And you also had a *collective* self-conception, i.e., a conception of who you were and what you were like, as a pair. You constructed this latter self-conception gradually over time, by compromising and coordinating with respect to

many practical issues, and then attributing a preference for these compromises to you and Doug, as a person. For instance, you value work more than play, and Doug values play more than work; and therefore you think of you and Doug, as a person, as valuing work and play in equal measure. Moreover, you often switched back and forth between your individual and collective self-conceptions many times everyday, in order to decide what to do. Specifically, you identified with your individual self-conception in order to decide what to do by yourself. i.e., you asked yourself, "What do I want to do right now, and how can I do it?" And you identified with your collective self-conception in order to decide what you and Doug should do together. i.e., instead of asking yourself, "What do I want to do, and what does Doug want to do, and how can I do this, and how can he do this?" you projected yourself into the standpoint of you and Doug, as a pair, and asked yourself, "What do we want to do, and how can we do it together?" This made it much easier for you and Doug to work together. And of course, as your fusion progressed, you found it useful to do this much more often, because you had to compromise and coordinate much more often, and with respect to many more issues (and, fortunately, you also had a much more detailed collective self-conception as a result).

None of this has changed at all. You still have an individual as well as a collective self-conception, and you still identify with both in everyday life. There are only two differences now. First, you think and talk about each self-conception differently. Instead of thinking about your collective self-conception as "us" and your

individual self-conceptions as "me" and "Doug", you think about your collective self-conception as "me" and your individual self-conceptions as "my day self" or "me right now," and "my night self" or "me tonight." Second, and as a result, your "default" self-conception changes: you now identify with your collective self-conception by default, and you can identify with your individual self-conception only with psychological effort.

How does this change your relationship with Doug? In other words, how does it change your relationship with your night self? In several ways. First, it makes it easier for you to get along. Insofar as you identify with a shared self-conception, you will make decisions from a shared practical standpoint. Thus, even if you continue to have different beliefs, desires, aims, and so on as personalities, you will still agree about what to do, since you will each think about what to do from a practical standpoint that represents a fair compromise between you. Thus, if you identify with your shared self-conception by default, you will also share intentions and actions by default. Indeed, you will be able to act on your own only with psychological effort, only by projecting yourselves back into your individual self-conceptions before deciding what to do.

Nevertheless, you can still do this if you want to. And insofar as you do, you will still experience the same kind of conflict as before, and will still have to use third-personal strategies for interacting with each other. For example, even though you agree with your night self that you should go on a trip, and even though you

agree about the general parameters of the trip, you also disagree about many of the specifics. For example, you really want to get an early start on the first day of the trip. (You will be indifferent about this tonight.) However, you know that getting an early start will be impossible if you drink tonight, since anytime you drink, it takes you forever to get out the door the next day. Unfortunately, you also know that when the night rolls around, you will want to drink. So what should you do?

Well, since you and your night self still disagree about whether to drink, you still have the same two options as before. First, you can try to compromise. For example, you can ask yourself not to drink tonight for your sake, and you can promise to let yourself have a few extra drinks tomorrow night in exchange. And, as always, if you really think that your offer is fair (and that you will agree tonight), then you can skip your negotiations entirely and express your offer in the form of an intention: you can intend not to drink at all tonight, and then to drink a lot tomorrow night, with the expectation that you will act on this intention later. Alternatively, if you want to get an early start tomorrow morning without having to give up anything in exchange, then you can accomplish this by attacking and threatening your night self. For example, you can take all your beer and pour it down the drain, take all your money and give it to a friend to keep until morning, and then lock yourself in a room until your friend returns. Then you can tell yourself that if you so much as think about drinking tonight, you will do everything in your power

to make sure that you stay sober the whole rest of the trip, and maybe for much longer than that.

The question is, as always: What does morality have to say about this? Are you morally permitted to do this to your night self on the grounds that you now think and talk differently and, as a result, identify with yourself as a person more often? Intuitively, the answer is still no. The fact that you now call your night self "my night self" instead of "Doug," and the fact that you now call your day self and night self "me" instead of "us," does not free you to do whatever you want to your night self. Similarly, the fact that you now identify with your collective self-conception more often, and therefore share intentions and actions more often, does not free you to do whatever you want to your night self when you identify with your individual self-conception instead. In short, then: you still have a duty to let your night self live the way he wants, and vice versa. Fortunately, insofar as you share a self-conception, you can do this by default, without even thinking about it. But insofar as you still identify with your individual self-conceptions (i.e., insofar as you still think of yourself as 'my day self' or as 'my night self'), you still have a duty to make sacrifices for each other, as well as a right to bind each other only insofar as you have to, in order to protect your rights against each other.

I have argued for two main conclusions in this thought experiment. First, you have a duty to let your night self live the way he wants, and vice versa, even if you share a

body, a brain, a very high degree of psychological connectedness, and a self-conception, and, therefore, even when you become parts of the same person, on most theories of personal identity.⁵⁵ Second, however, you also have a right to bind each other more than you did before, since you have a more intimate relationship than you did before, and therefore you may *have* to bind each other more now, in order to protect your own rights and continue to live and let live.

I believe that our intuitions in support of these conclusions come from our recognition of an important truth about morality: we have moral duties to each other and rights against each other not because we have different bodies, different streams of consciousness, or different self-conceptions, but rather because we all have ends to pursue, and sometimes these ends may conflict. Insofar as our ends are compatible, great: instrumental rationality will direct us to work together to promote shared aims, and to otherwise keep to ourselves. But insofar as our ends conflict,

⁵⁵ Specifically, you and Doug are the same person on any theory that makes personal identity a matter of sharing a body, sharing a brain, or sharing a stream of psychological connectedness and/or continuity. The only possible exceptions are non-reductionist theories, which make personal identity a matter of a “further fact,” like having a soul, and self-narrative theories, which make personal identity a matter of telling a particular self-narrative across time. But in my view, non-reductionist theories are implausible, and as I argue in “Practical Integrity,” many people actually have *multiple* self-narratives: we have a self-narrative for each personality, as well as another, more general self-narrative for the person as a whole, which we construct, as Dennett puts it, by finding the “center of narrative gravity” in our thoughts, feelings and actions across contexts. I think that the fact that we have multiple self-narratives makes them a poor criterion for personal identity (as do other features of self-narratives). But I also think that sharing self-narratives with our future selves helps us to share intentions and actions with them easily and effectively; and therefore I think that self-narratives, if not a good criterion for personal identity, are at least a good (though imperfect) criterion for *collective agency*.

instrumental rationality will direct us to take the most expedient route to promoting our own ends. This is when morality steps in: whenever the most expedient route to promoting our own ends involves trampling over others in the process, morality draws us back and instructs us to take a different route instead.⁵⁶ And this is true whether the ‘others’ in question are other people, on one hand, or other personalities, on the other. In fact, if anything, I have argued that the limitations that morality places on instrumental rationality are all the more *important* when the ‘others’ in question are other personalities, since our relationships with them will affect many more aspects of our lives than our relationships with other people will.⁵⁷

⁵⁶ Herman 1993 makes the same point in more explicitly Kantian terms:

Manipulative interventions in the process of willing (deception, coercion) that regard the will as just another route towards ends are ... impermissible. Such actions do not respect the integrity of the will (the maxims of such manipulative interventions cannot have the form of universal lawgiving): they accord the will only relative value. (p. 156)

⁵⁷ One might agree with me that you still have duties to Doug at the end of fusion, but object to my calling these duties ‘moral,’ since one might think that moral duties are by definition *interpersonal*. I have two things to say about this. First, my interest is in what to do, not how to talk. So as long as you agree with me about how you should treat Doug, I am happy for you to classify this duty however you want. Second, however, I think that some ways of classifying duties make more sense than others. Specifically, I think it makes more sense to classify duties based on why we have them, rather than (our pre-theoretical intuitions about) whom we have them to. And in this case, you and Doug have a duty to let each other live the way that you want as *personalities* for the same reason that you did as *people*, namely, you each have ends to pursue, and sometimes these ends may conflict, and therefore you each have a duty to pursue your ends in a way that lets the other do the same. And if this is right, then what reason do we have to call your duty to Doug-the-person ‘moral’ and your duty to Doug-the-personality ‘non-moral’? It seems much simpler to just call them both ‘moral’. But as I said, as long as you agree with me about what to do, then you can carve up the terminological terrain however you like.

2.2. Multiple Personalities

This is all well and good, but does it actually apply to any of us? I think that it very well might. Keep in mind that I mean ‘personality’ here in a very minimal sense, according to which a personality is just a characteristic set of beliefs, desires, aims, and so on. So my conclusions apply not only to people with multiple personality *disorder* (i.e., people with multiple *psychologically disconnected and discontinuous* personalities), but also to people with multiple personalities in the mundane, non-pathological sense of having a “day self” and a “night self,” or perhaps more realistically, a “work self,” a “home self,” a “friend self,” and so on (i.e., people with multiple *psychologically connected and continuous* personalities).

In my view – and in the view of many contemporary personality psychologists⁵⁸ – many people have multiple personalities in this non-pathological

⁵⁸ The idea that many people have multiple personalities in a non-pathological sense is relevant to the debate between “dispositionism” and “situationism” in personality psychology. Dispositionists believe that we should explain our behavior by reference to internal character traits, whereas situationists believe that we should explain our behavior by reference to external features of the environment. In contrast, I accept the middle-ground, “interactionist” view that we should explain our behavior by reference to both (a) internal character traits and (b) the external features of our environment that trigger those character traits. Fortunately, many personality psychologists now seem to think that this middle ground, interactionist view is basically correct, and that the only questions are, e.g., how many personalities we have, how distinct they are, how sensitive to environmental triggers they are, and so

sense.⁵⁹ We might not think that we do. We might think that we just play different social roles. But actually, as I argued in chapter 1, I think that our playing different social roles is part of what *makes* us have different personalities. Specifically, in the course of mastering each social role, we begin to take on the traits of our character in that role, like method actors, until eventually we experience a kind of partial psychological fission: we develop a different personality for each role that we play, and then, when we enter the relevant context, the relevant personality comes out naturally.⁶⁰ Granted, we might undergo only minor psychological changes when this happens: we might just find that a few thoughts, feelings, and habits become active, a few others become dormant, and a few others become more or less salient than they were before. But even minor psychological changes like this can have a major impact on our overall character and behavior. For instance, when you come home from the office – when you walk in the door, or change out of your work clothes, or sit down at the dinner table – you might, as a result of these minor psychological changes, find that you feel like a different person altogether (or as you might put it, you might find

on. See Doris (2005), Harman (2000), and Upton (forthcoming) for discussion about how this debate affects virtue ethics.

⁵⁹ Bloom, Carter, and Hacking use ‘personality’ in this minimal sense as well.

⁶⁰ If this is right, then why do many of us seem so unified? There are many reasons for this, but one of the main reasons, in my view, is: even though we have multiple personalities, we also have a person-level self-narrative, which we construct by taking the “center of narrative gravity” in our thoughts, feelings, and habits across contexts. We then identify with this person-level self-narrative more often than with our personality-level self-narratives. And this makes us seem, internally, to be less fragmented than we actually are. (Happily, this also has the effect that we often think and act from a perspective that represents a fair compromise among our personalities.) See my “Practical Integrity” for more on this.

that a different “part” or “side” of you comes out). You stand with a different posture, you speak with a different tone of voice, you have a different mood, and most importantly for our purposes here, you have a different perspective on what really matters in life – or at least, you change your mind, in certain respects (large or small), about what to do.

This kind of normal, non-pathological psychological change is all it takes for you to be the kind of person that I have in mind here. For example, suppose that you are forced to decide, at work, whether to take a promotion that will require you to spend more time at work and less time at home. My argument implies that, even if you think, right now, that your career is your priority, and even if you would be willing, right now, to spend less time with your family in order to advance in your career, you have a moral duty to take your “home self” into account when making your decision too. Or to put the point in more general terms: whenever you have a decision to make that will affect your other personalities as well (as almost all of your decisions will), you have a moral duty to act on behalf of the person as a whole, i.e., you have a moral duty to try to do what “we” prefer to do instead of just what “I” prefer to do.⁶¹ Not only because this will make your other personalities more likely to accept your decision, but also because they have a right to live the way they want just as much as you do.

⁶¹ Of course, you may use the first-person singular instead of the first-person plural to refer to all of your personalities. But this is normatively irrelevant.

Of course, it may not always be possible for you to make a decision that makes all of your personalities equally happy. For instance, some personalities may pose more of a threat to the “general population” than others, and therefore, any fair compromise will probably make them less happy. Or take a more extreme example: two personalities may want different and radically incompatible futures – e.g., one may want to get married and another may want to stay single, or one may want to move to Texas and another may want to stay in New York – and thus, any choice that you make will leave at least one of these personalities very unhappy. In fact, your choice may have even more significance than that: if, as I believe, your personalities are context-dependent, then your choice may determine which personality gets to come out at all, or at least, which one gets to come out more frequently: e.g., if you decide to get married, then your “single” personality may come out much less often, or if you decide to move to Texas, then your “New York” personality may come out much less often. In a sense, then, some of your most important decisions in life may be like lifeboat cases – cases where no matter what you do, only one personality, or group of personalities, will get to survive and flourish.⁶² As in lifeboat cases, certain personalities may be justified in trying to be the ones who survive, and may also be justified in binding or harming their fellow

⁶² I mean ‘survive’ metaphorically, since as alcoholics know all too well, avoiding a certain context may keep a certain personality dormant, but it will rarely kill it altogether. Perhaps, then, a more accurate metaphor would be: any choice that you make will cause some personalities to get locked away in a high-security psychological prison.

personalities in order to do so. But also as in lifeboat cases, not all personalities will be justified in doing this; and either way, they should regard the situation in general as a moral tragedy (though perhaps on a smaller scale than in an actual lifeboat case).

Of course, this is only part of the story about self-binding. You might be justified in binding your other personalities on many other grounds too. For example: some personalities might be mistaken about what they want or how to get it, and therefore you might be justified in binding them on paternalistic grounds. Moreover, some personalities might not even be *capable* of compromising with you. It all depends on what kind of personality they are, and when they come into existence. For instance, some personalities are like animals, in the sense that they have first-order desires but no ability to step back from them and ask whether to act on them, much less compromise with *you* about this. (Many addictive and compulsive personalities are like this.) Similarly, some personalities are like future persons, in the sense that they are born after you die, and therefore you have no opportunity to jointly commit to a mutually acceptable way of life and hold each other to the promises you make. (Parfit's "Nineteenth Century Russian" case is like this.⁶³) And as I argue elsewhere, I think that your inability to compromise with these personalities,

⁶³ "In several years, a young Russian will inherit vast estates. Because he has socialist ideals, he intends, now, to give the land to the peasants. But he knows that in time his ideals may fade." (1984, p. 327).

along with other features of your relationship with them, justifies your interfering with them more than with personalities like Doug.⁶⁴

Taken together, these considerations will make it the case that self-binding is often, though not always, morally permissible all things considered. They will also make it the case that you can permissibly bind irrational, compulsive, and destructive personalities more often than their opposites. I think that this demonstrates that, even if my view is revisionary in some respects (since it implies that some cases of self-binding are morally wrong), it is ultimately intuitively plausible as well (since it implies that most “standard” cases of self-binding are morally permissible all things considered). But we have to remember two things: first, you have to have a very good reason in order to justify binding your other personalities. And second, this has to be the same kind of reason that would justify your binding other people too.⁶⁵

⁶⁴ See my “The Personal Community.”

⁶⁵ One might wonder how my view applies to people with just one personality. The answer is: If you have just one personality, then moral questions about how to treat your future selves never arise. You always have the same beliefs, desires, aims, and so on across contexts, so instrumental rationality will always recommend working with your future selves to promote shared aims, rather than binding or otherwise harming them.

Ultimately, then, what matters here is *not* whether you have the same personality as your future self, but rather whether you agree with him about what to do. On my understanding of ‘personality,’ these go hand in hand, since you have the same personality as a future self to the degree that you have the same beliefs, desires, aims, etc. as him. So the more you share a personality with a future self, the more likely you are to agree with him about what to do. And I suspect that, in this sense of ‘personality,’ even very unified people will have *partially* multiple personalities, since they will experience minor psychological changes that lead them to – infrequently, and in minor ways – disagree with their future selves about what to do.

2.3. The Just Soul

I think this discussion supports a striking conclusion, namely, many people are like *nations*, in the sense that they are *communities* of moral agents (i.e. moral duties- and rights-holders) who share resources and a common fate, and who therefore have a duty to jointly commit to a fair set of laws, as well as a right to bind each other only insofar as they have to, in order to enforce these laws.

In fact, we might have started even further back in our thought experiment, in order to drive this point home. For instance, we might have started by making you and Doug members of the same international community, then members of the same nation, then members of the same state, then members of the same community, then members of the same household, and then members of the same person. And each step of the way, we would have seen the same kind of change take place. You would still have the same kind of moral relationship that you had before, but you would also share territory and resources in a more intimate way, and therefore you would have a duty to establish a new, more expansive set of laws meant to govern your treatment of each other.

Of course, there are a number of important differences between a nation and a person in addition to how expansive its set of laws should be. For example, a nation is so large and diffuse that, practically speaking, its citizens will have to compromise and coordinate with each other *indirectly*, by participating in political institutions,

rather than *directly*, by talking through their problems or forming and acting on intentions in the spirit of compromise. But as I have argued at length, none of this is morally relevant for present purposes. Our methods of interaction – our methods of writing, interpreting, and enforcing the laws that we make together – might change as our group gets smaller and more intimate, but the goal is always the same: to find a set of policies that will help us to live and let live.

Thus, I think that Plato was right: we should use political morality as a model for personal morality. Of course, I arrived at this conclusion in a very different way than Plato did. Plato argued that the soul is divided into a rational part, a spirited part, and an appetitive part. Thus, he claimed, the soul resembles the state, which is divided into “lovers of wisdom,” “lovers of honor,” and “lovers of money.” He then argued that, if this is right, then the just soul, like the just state, is a *dictatorship*, where the rational part rules the other parts. In contrast, I have argued that the soul is, or at least can be, divided into personalities (many of which are rational, spirited *and* appetitive). Thus, the soul resembles, or at least can resemble, the state, which is divided into people. And I have suggested, though not yet in these terms (but see below), that if this is right, then the just soul, like the just state, is a *liberal democracy*, where each personality has a right to live the way that he wants, as well as a right to participate in collective decision-making.⁶⁶

⁶⁶ This point requires some finesse, since not all personalities will be fully rational and capable of participating in collective-decision making; and in these cases, it might be better to have other personalities make decisions on their behalf. Of course, the

The challenge, then, will be to apply our best theories of political morality to people as well as to other, larger collectives, while being mindful of the fact that, in light of the many differences between people and other, larger collectives, we may need to implement our theories very differently at each level. In what follows, I will briefly sketch how we might do this with respect to two central concepts in political morality: justice and sovereignty. To be clear, I will not be arguing for any particular theory of justice or sovereignty here (though the theories that I pick are plausible and widely accepted, and to some degree supported by my arguments in this paper). I will rather just be illustrating my claim that we can, and should, use resources from political philosophy to shed new light on what we owe to ourselves, and to each other.

First, consider how we think about the just state. As Rawls argues, the concept 'justice' applies to a group of people if and only if the *circumstances* of justice apply to that group. What are the circumstances of justice? They are the circumstances in which this virtue is possible to have, as well as necessary to try to have. So what are they in particular? Rawls provides a long list, which includes the following: the members of this group have to have different conceptions of the good, or in other words, different ends in life. They have to live in the same territory and use the same resources, often under conditions of moderate scarcity. They have to have roughly similar mental and physical powers, and be motivated partly by self-

same is true about people. Thus, this is just another respect in which the soul is like the state.

interest and partly by benevolence. Finally, they have to be vulnerable to attack, both from each other and also, as a group, from other groups.⁶⁷ Rawls then claims that if these circumstances apply to a group of people, then they can, and must, find a way to live together in a just state. He also claims that the best way for them to do this is not by *eliminating* the circumstances of justice, but rather by *responding* to them in the right kind of way⁶⁸: by jointly committing to a fair set of laws so that they can get along with each other, as well as so that they can get along, as a state, with other states.

How does this theory of justice apply to the soul? I have argued, though not in these terms, that the circumstances of justice apply to groups of *personalities* just as much as, if not more than, to groups of *people*. For instance, your “work self,” “home self,” “friend self,” and so on have different conceptions of the good, and therefore they also have different ends in life. (They may agree about how to live very often, and with respect to very many issues. But they still disagree about how to live sometimes, and with respect to some issues, and this is enough for them to satisfy this condition.) They also live in the same territory (i.e., not just in the same nation, state, and household, but also in the same body, brain, and mind) and they use the same resources, often under conditions of moderate scarcity (i.e., not just resources like food, water, and electricity, but also “personal resources” like health, strength,

⁶⁷ Rawls, 1999, 109-12.

⁶⁸ As Rawls puts it: “Unless these circumstances existed there would be no occasion for the virtue of justice, just as in the absence of threats of injury to life and limb there would be no occasion for physical courage” (1999, p. 110).

and energy). Moreover, they are roughly similar in physical and mental powers, and motivated partly by self-interest and partly by benevolence. Finally, they are vulnerable to attack, both from each other and also from other people. Therefore they can, and must, try to live together in a just soul. And as Rawls argues (and as I have argued), the best way for them to do this is not by *eliminating* the circumstances of justice – for example by conditioning themselves to have the same ends so that conflict between them never arises (after all, they might have good reason to want to stay psychologically fragmented to an extent⁶⁹) – but rather by *responding* to the circumstances of justice in the right kind of way: by jointly committing to a fair set of laws so that they can get along with each other, as well as so that they can get along, as a person, with other people.⁷⁰

⁶⁹ I argue for this conclusion in “Practical Integrity.”

⁷⁰ I read Korsgaard as making a similar point when she writes

The requirements for unifying your agency internally are the same as the requirements for unifying your agency with that of others. Constituting your own agency is a matter of choosing only those reasons you can share with yourself. . . . This is what Plato means – and he means it literally – when he says that a good person is his own friend, and legislates for the good of his soul as a whole. (2009, pp. 202-3)

What does it mean for you to “unify your agency” through “self-constitution”? Not that you should try to have the same beliefs, desires, aims, etc. across time. This is not how you unify your agency with that of others, and Korsgaard never suggests that you should do it with respect to yourself. Instead, I think Korsgaard is saying that you should unify your agency in a *practical* sense rather than in a *psychological* sense: you should act on principles that you can endorse from within all of your practical identities, even you have different beliefs, desires, aims, etc. from within these identities. In other words, you should ratify a literal self-constitution – a set of laws meant to govern your treatment of yourself, so that you can live with yourself, and with others, peacefully and respectfully. If you do this, then you will be acting on

Second, consider how we think about national sovereignty. Many philosophers think that, at this level of application, sovereignty has an intimate connection with justice. The reason is that sovereignty applies not only to *nations* but also to their *citizens*, and therefore we need a theory of sovereignty that works at both levels at once. And this is exactly what many contemporary political philosophers have set out to develop.⁷¹ Roughly speaking, we think that each nation has the right to set its own laws, free from the interference of other nations, provided that it respects their right to do the same. Does this mean that a nation's citizens are morally permitted to treat each other however they want, or that a nation's public officials are morally permitted to write, interpret, and enforce the law however they want? Of course not. On the contrary, we think that a nation's citizens have a duty to live and let live, and their public officials have a duty to write, interpret, and enforce the law in a way that helps them to do this. So what does it mean for a nation to have sovereignty, then? It means this: If a nation's citizens create a just state together – if

“reasons you can share with yourself” and legislating “for the good of [your] soul as a whole,” rather than just your part of it.

⁷¹ Generally speaking, there have been two major trends in the literature on national sovereignty since World War II. The first is that we began to set more limits on what nations can do to *each other*, e.g., we now believe that nations have a duty not to attack each other unless they have to, in order to defend itself or other nations. Second, we began to set more limits on what nations can do to *themselves*, e.g. we now believe that a nation's public officials have a duty not to write, interpret, or enforce the law in a way that seriously harms or exploits their fellow citizens. One way of thinking about the connection between national sovereignty and personal sovereignty is that, with respect to the former trend, our theories of personal sovereignty have been ahead of our theories of national sovereignty; but with respect to the latter trend, our theories of national sovereignty have been ahead of our theories of personal sovereignty. Thanks to Sam Scheffler for pointing this out to me.

they jointly commit to a fair set of laws – then they have a collective right to live under these laws; or in other words, the *nation*, understood as a collective agent⁷², has a right to live under these laws, free from the interference of other nations. And importantly, this is true even if other nations disagree with it about which laws to establish, and even if the laws that it establishes bind or harm some of its citizens in various ways⁷³. However, if a nation’s citizens fail to create a just state together – if

⁷² Depending on your theory of collective agency, you may or may not think that a nation is *itself* a collective agent. For example, if you think (as I do) that people need to share intentions and actions in order to be a collective agent, then you will think that a nation only roughly corresponds to a collective agent, because while many of its citizens will participate in the relevant collective actions, some of its citizens will not, and other people who are not citizens will. Nevertheless, other nations may have a right to *treat* a nation as a collective agent for certain practical purposes. See my “Agency Is Collective All The Way Down” for more on this.

⁷³ It is a controversial question what, if anything, gives the state the right to bind its citizens, and what, if anything, gives individual citizens a duty to follow the law. These two questions are known as the problem of political authority (or legitimacy) and the problem of political obligation, respectively. Without wading too deeply into these issues here, let me just say three things.

First, my claims in this paper are most in line with Margaret Gilbert’s joint commitment theory of political obligation. Gilbert argues that citizens have a duty to follow the law if “they are jointly committed to uphold as a body a particular set of institutions of governance” (2006, p. 288). Importantly, she also thinks that citizens can jointly commit to upholding the relevant institutions without actually or tacitly agreeing to do so, which is why her theory avoids the problems of actual or tacit consent theory.

Second, even if this theory failed at the national level, it might still succeed at the personal level. After all, a person’s personalities have a much more intimate relationship than a nation’s citizens do. Thus it is much easier to imagine how they might make the relevant joint commitments, as well as why they might be motivated to do so.

Third, some philosophers think that political authority implies political obligation, whereas others (including me) deny this. On this latter view, a nation’s public officials can be permitted to enforce the law whether or not their fellow

they collapse into civil war, or if they jointly commit to an unfair set of laws – then other nations are permitted to intervene in order to stop them from hurting each other (as well as, of course, in order to stop them from hurting the citizens of other nations). Indeed, if a nation is at war with itself – if its citizens fail to jointly commit to any set of laws at all – then it is not a collective agent with any interests or rights to begin with.

How does this theory of sovereignty apply to the soul? We can start by saying that, at this level of application as well, sovereignty has an intimate connection with justice. The reason is that sovereignty applies not only to *people* but also to their *personalities*, and therefore we need a theory of sovereignty that works at both levels at once. How might such a theory look? Based on our analogy, we can say that it might look like this: a person has a right to live the way that he wants, free from the interference of other people, provided that he respects their right to do the same. Does this mean that a person's personalities are permitted to treat each other however they want? No. They have a duty to live and let live too. So what does it mean for a person to have sovereignty then? It means this: if a person's personalities create a just soul together – if they jointly commit to a fair set of laws – then they have a collective right to live under these laws. In other words, the *person*, understood as a collective agent, has a right to live under these laws, free from the interference of other people. And this is true even if other people disagree with him

citizens are permitted to follow it. And if this is true about a nation's citizens, then it is also true, I believe, about a person's personalities.

about how to live, and even if his way of life binds or harms some of his personalities in various ways. However, if a person's personalities fail to create a just soul together – if they collapse into civil war, or if they jointly commit to an unfair set of laws – then other people are permitted to intervene in order to stop them from hurting each other (as well as, of course, in order to stop them from hurting personalities in other people). Indeed, if a person is at war with himself – if his personalities fail to jointly commit to any way of life at all – then, we might say, he is not a collective agent with any interests or rights to begin with.⁷⁴

To repeat, this is just a sketch of a theory of personal morality based on political morality. We still have many more questions to answer before we have anything like a full theory. Nevertheless, we can draw two preliminary conclusions. First, whichever political theory we ultimately accept, we should be prepared to apply it to *people* as well as to other, larger collectives – and we should take it as a constraint on our theory selection that it has plausible implications at each of these levels.⁷⁵ And

⁷⁴ I argue that we should understand persons as collective agents in “Agency Is Collective All the Way Down.”

⁷⁵ That said, we might have reason to place more weight on our intuitions about the just state and national sovereignty than on our intuitions about the just soul and personal sovereignty, since, as Plato pointed out, nations are larger and therefore easier to think clearly about. As Socrates puts the point in a conversation with Adeimantus:

“We say, don’t we, that there is the justice of a single man and also the justice of a whole city?”

“Certainly.”

“And a city is larger than a single man?”

second, if we apply the kind of political theory that I have sketched here to people (and I should emphasize that many of us accept this kind of theory), then the result is that personal sovereignty is harder to achieve than many philosophers have thought. Specifically, you have to satisfy at least two conditions in order to enjoy the right to live the way that you want: first, your personalities have to make joint commitments in a way that makes you, as a person, into a collective agent. And second, these joint commitments have to be *just*, i.e., they have to be fair to your personalities, as well as to other people.

2.4. Conclusion

I have argued for three main conclusions in this paper. First, self-binding is *prima facie* morally wrong. However, you can still bind “fellow personalities” more than “foreign personalities” for the same reason that you can bind fellow citizens more than foreigners, namely, you may *have* to bind them more in order to protect your own rights.

“It is larger.”

“Perhaps, then, there is more justice in the larger thing, and it will be easier to learn what it is. So, if you’re willing, let’s first find out what sort of thing justice is in a city and afterwards look for it in an individual, observing the ways in which the smaller is similar to the larger.” (*Republic II* 368e2-369a2).

Second, many people are like nations, in that they are *communities* of moral agents who share resources and a common fate. Thus, Plato was right that we should use political morality as a model for personal morality.

Finally, if we apply a plausible and widely accepted theory of political morality to the person, then we get the following results: a person has a just soul insofar as, and only insofar as, his personalities jointly commit to a fair set of laws, and a person has a right to live the way that he wants insofar as, and only insofar as, he has a just soul.

3. Personal Responsibility

“What happens in Vegas stays in Vegas”

- The City of Las Vegas

I have argued that if a person has multiple personalities, then these personalities are separate agents, i.e., separate bearers of reasons, duties, and rights. Moreover, I have argued, this is true even if these personalities share a chain of memory, a very high degree of psychological connectedness, and a self-conception. For example, you might have a “family self” that comes out around your family, a “work self” that comes out around your colleagues, a “friend self” that comes out around your friends, and so on, where the difference between these “selves” is just that each tends to have a few beliefs, desires, habits, and so on that the others lack (or even just have less strongly). For instance, you might think, around your colleagues, that you should take a promotion that will require you to spend more time at the office, whereas you might think, around your family, that you should turn it down. Similarly, you might think, around your friends, that you should stay out and have a few extra drinks, whereas you might think, around your family, that you should come home in time for

dinner. If so, then, I have argued, each personality has a right to live the way that it wants, as well as a duty to let the others do the same. Thus, they have a duty to jointly commit to a fair set of “personal laws,” as well as a right to bind each other only insofar as they have to in, order to enforce those laws.

My goal in this chapter is to explore the implications of this idea for personal responsibility. Specifically, if each personality is a separate agent, and if each agent is praise- or blameworthy only for what it does,⁷⁶ then each *personality* is praise- or blameworthy only for what it does. Thus, for example, if your work personality decides to take the promotion, then your work personality might be blameworthy for this decision, but your other personalities are not. Similarly, if your friend personality decides to stay out and has a few extra drinks, then your friend personality might be blameworthy for this decision, but your other personalities are not. Thus, when you wake up the next morning and your family angrily demands, “Why did you take that job without consulting us?” or “Why did you stay out all night with your friends?” You can truly reply, “A different personality did that,” with the implication that the personality who performed this action deserves blame for it, not the personality currently in charge. And of course, these are relatively minor “crimes”: the same conclusion follows for more serious crimes too. For example, if a new personality

⁷⁶ Some philosophers believe that we should reject the idea of praise- or blameworthiness entirely. I think this view is plausible. Nevertheless, I will assume for the sake of argument that agents are praise- or blameworthy only for what they do, since my main purpose here is to show that my conception of agency is no threat to those who hold this view, and who want to be able to justify many of our everyday judgments about personal responsibility.

comes out on your trip to Vegas and has an affair, or robs a casino, or drives a truck into an outdoor cafe, then your other personalities will not be blameworthy for any of these actions when they come back home. Thus, on my conception of agency, there really is a sense in which “what happens in Vegas stay in Vegas.”

Many philosophers, of course, would reject these implications. They believe that, setting aside cases of amnesia, mind-control, and the like, people are praise- or blameworthy for what they did in the past, even if another personality was in charge when they did it. As a result, many philosophers accept much “stronger” conceptions of agency than mine, according to which your present self is part of the same agent as a past or future self if they, say, (a) share a body/brain, (b) share a chain of memory, (c) share a high degree of psychological connectedness, or (d) share a self-conception.⁷⁷ These conceptions of agency, together with the assumption that an agent is praise- or blameworthy for his actions, all imply that, if you have multiple personalities in a non-pathological sense, then your personalities are part of the same agent, and this agent is praise or blameworthy for everything that it does. Thus, for example, you are blameworthy for taking the promotion, staying out late and having

⁷⁷ Not all philosophers offer these theories *as* theories of moral agency in the first instance. Some offer them as theories of personal identity in the first instance, and then add that my present self is part of the same *moral agent* as a past or future self iff he is part of the same *person* as a past or future self. I will use the term ‘person’ in a way that makes people come apart from moral agents. Specifically, I will assume that if your present self shares a body/brain, a high degree of psychological connectedness, and a self-conception with a past or future self, then they are part of the same person, but not necessarily part of the same moral agent. But I do not expect anything to turn on this terminological choice.

a few extra drinks, driving the truck into the outdoor café, and so on, not only at the time, but also when you wake up the next morning and wonder “what the hell did I just do?”

I believe, however, that we can justify many of our everyday judgments about moral responsibility *without* accepting such a “strong” conception of agency. The key is to see that we use the term ‘moral responsibility’ in many different senses, several of which are compatible with the idea that one agent can be ‘morally responsible’ for an action *without* being praise- or blameworthy for it. For example, we can be *collectively* responsible for what we do together, we can be *indirectly* responsible for what other agents do, we can be *criticizable* in light of what other agents do, and we can be *liable* for what other agents do. And in my view, a conception of agency that attributes praise- and blameworthiness to us sparingly, and then *supplements* this with attributions of these other kinds of moral responsibility, ultimately justifies our everyday judgments about moral responsibility – including our judgment that there really is, in some cases, a weak sense in which “what happens in Vegas stay in Vegas” – at least as well as a conception of agency that attributes praise- and blameworthiness to us very liberally.

I will proceed as follows. In section 1, I will consider several standard senses of ‘moral responsibility,’ and show how, in each of these senses, we can be ‘morally responsible’ for an action without having performed it, and without being praise- or blameworthy for it. Then, in section 2, I will argue that, if we take into account all

these different senses of ‘moral responsibility,’ we will see that my conception of agency accommodates our judgments about personal responsibility at least as well as other, more traditional conceptions.

3.1. Different Senses of ‘Responsibility’

Assume for the sake of argument that, if a person has multiple personalities in a non-pathological sense, then each personality is praise- or blameworthy only for what he does. There are still many senses of the term ‘moral responsibility’ in which each personality can be morally responsible for what others do (or in light of what others do, or as a result of what others do). In this section, I will present several (though there are undoubtedly more).

Collective Responsibility

First, even if personalities are separate agents, they can still be part of the same *collective* agent. Correspondingly, even if each personality is praise- or blameworthy only for what he does, they can still be *collectively* praise- or blameworthy for what they do together.

What does it take to be part of the same collective agent? There are many answers to this question in the literature. For purposes of my discussion here, I will focus on a standard theory of collective agency, the *shared intention* theory, as developed by Michael Bratman.⁷⁸ Here is how this theory works. Suppose that a group of agents build a house together. Do they count as a collective agent? That depends on why they do what they do. If each is just acting completely independently of the others, but their efforts just happen to result in a house being built, then they do not count as a collective agent. But if (as is much more likely) each does what he does at least in part because he *intends* for them build a house together, then they count as a collective agent, and they also, as a result, count as collectively praise- or blameworthy for what they do together.

It is worth noting how flexible this theory of collective agency is. For example, a group of agents can share an intention to build a house even if they each act at different *times*, even if they each act in different *ways*, even if they each have different *reasons* for sharing this intention, and even if they are not consciously aware

⁷⁸ See, for example, the many essays collected in Bratman, 1999, *Faces of Intention* (New York: Cambridge University Press, 1999). For alternative theories of shared intention, see Gilbert, Margaret. 1990, "Walking Together: A Paradigmatic Social Phenomenon," *Midwest Studies in Philosophy* 15:1–14; Kutz, 2000, *Complicity: Ethics and Law for a Collective Age* (Cambridge: Cambridge University Press); and Velleman, 1997, "How to Share an Intention", *Philosophy and Phenomenological Research* 57:29–50. Though I am focusing on Bratman's theory here, I believe that all of these other theories imply that personalities can share intentions too.

of sharing this intention at all.⁷⁹ For example, a group of agents can share an intention to build a house even if one of them finances the project, another buys the materials, another builds the foundation, another puts up the walls, and so on. Further, they can share this intention even if one of them is doing it for the money, another is doing it for the exercise, and so on. Finally, they can share this intention even if they are not consciously aware of sharing it at all: as long as the best explanation of their behavior is that each of them intends, among other things, that they build a house together, they count as sharing an intention to build a house together, whether they know it or not.

How does this apply to personalities? Straightforwardly. Suppose that your personalities decide to build a house together. Do you, as a person (i.e. a group of personalities), count as a collective agent? That depends on why your personalities

⁷⁹ This last point – that agents can share an intention even if they are not consciously aware of doing so – is more controversial than others. In some papers, Bratman distances himself from this idea by claiming that, in “basic” cases of shared intention, there is “common knowledge” of it among those who share it. However, he does not argue that this is a necessary condition for shared intention. (See, for example, Bratman, “Shared Agency,” *Philosophy of the Social Sciences: Philosophical Theory and Scientific Practice*, C. Mantzavinos, ed., New York: Cambridge University Press, 41–59.) Meanwhile, other philosophers argue that common knowledge is *not* a necessary condition for shared intention. For example, Christopher Kutz argues that, on a functionalist account of intention, an agent intends to perform that action just in case this intention both causally and teleologically explains the their behavior. Thus, on a functionalist view of intention, two or more agents *share* an intention to perform an action just in case this shared intention both causally and teleologically explains *each* of their behavior. (See Kutz, 2000, *Complicity: Ethics and Law for a Collective Age*, Cambridge: Cambridge University Press, pp. 72-74.) I will assume here, along with Kutz, that common knowledge is not a necessary condition for shared intention. That said, if we reject this assumption, then personalities will still share intentions very often – just not as often as I suggest here.

do what they do. If each is just acting completely independently of the others, but their efforts just happen to result in a house being built, then you do not count as a collective agent. But if (as is much more likely) each personality does what he does at least in part because he *intends* for them build a house together, then you, as a person, count as a collective agent, and you also, as a result, count as collectively praise- or blameworthy for what your personalities did together. And note that, as before, this is true even if your personalities each acted at different *times*, even if they each acted in different *ways*, even if they each had different *reasons* for sharing this intention, and even if they were not consciously aware of sharing this intention at all.

If a person with multiple personalities is collectively praise- or blameworthy for what their personalities do together, does this mean that each personality is praise- or blameworthy for what the others do? For instance, if you, as a person, are collectively blameworthy for having an affair, does this mean that your “family personality” (who participates by lying to your family) is blameworthy for, say, sleeping with another person? This depends on whether and how collective praise- and blameworthiness *divides*. Some philosophers believe that it does not divide: individual agents deserve praise or blame for what they do, and collective agents deserve praise or blame for what they do, period. Others believe that it *does* divide: individual agents deserve praise or blame for what they do, collective agents deserve praise or blame for what they do, and individual agents are at least *complicit* in what the other members of their collective do, where this means that they deserve at least

some praise or blame for what the other members do. I will not take a stand here on whether collective praise- and blameworthiness divides. Instead, I will just note that, if it does, then this will justify not only the claim that you, as a person, are blameworthy for having an affair, but also the claim that your family personality is at least somewhat blameworthy for sleeping with other people.

Indirect Responsibility

Second, many philosophers believe that we can be praise- or blameworthy for the *foreseeable consequences* of our actions.⁸⁰ Thus, if a foreseeable consequence of one of your actions is that *someone else* acts wrongly, then you can be blameworthy for foreseeably causing or allowing them to act this way.⁸¹

For instance, suppose that you hand me a gun, or even let me pick up a gun, with the foreseeable consequence that I will kill somebody. Now suppose that, as a matter of fact, I kill somebody. In this case, many philosophers would say that you are blameworthy for what you did. This does not mean, of course, that you are

⁸⁰ For discussion of indirect responsibility in morality and the law, see Feinberg, 1984, *Harm to Others* (New York: Oxford University Press), and Hart and Honore, 1985, *Causation in the Law*, Rev. 2nd ed. (New York: Oxford University Press).

⁸¹ Some philosophers believe that we are praise- or blameworthy for *all* the foreseeable consequences of our actions, whereas others believe that we are praise- or blameworthy for only *some* of them. In order to sidestep this issue, I will restrict my focus here to cases that most people agree about.

blameworthy for killing someone (I am blameworthy for doing that), nor does it mean that you are just as blameworthy as I am (foreseeably causing or allowing me to kill someone does not warrant as much blame as killing someone does, at least not in typical cases). Nevertheless, your action in this case warrants at least some blame. This is especially true if you perform it intentionally, i.e., if you hand me the gun or let me pick it up, fully expecting me to kill somebody as a result. But it is also true if you do it negligently, i.e., if you hand me the gun or let me pick it up never suspecting that I might kill somebody as a result, even though it is clear that I might. (This is of course the point of accusations of negligence: we accuse people of negligence when they perform actions whose harmful consequences are foreseeable, yet not foreseen.)

How does this apply to personalities? In the same way. For instance, suppose that your “friend personality” brings a gun into the nightclub, or even leaves the gun in a nearby car, with the foreseeable consequence that your hot-blooded, impulsive “nightclub personality” will kill somebody as a result. Now suppose that, as a matter of fact, your hot-blooded, impulsive nightclub personality kills somebody. In this case, your friend personality is blameworthy for what he did. This does not mean that he is blameworthy for killing someone (your nightclub personality is blameworthy for doing that), nor does it mean that he is just as blameworthy as your nightclub personality is (foreseeably causing or allowing a different personality to kill somebody does not warrant as much blame as killing someone does, at least not

in typical cases). Nevertheless, your friend personality's action in this case warrants at least some blame. This is especially true if he performs it intentionally, i.e., if he brings a gun into the nightclub fully expecting that your nightclub personality will shoot somebody as a result. But it is also true if he performs it negligently, i.e., if he brings a gun into the nightclub never suspecting that your nightclub personality might shoot somebody as a result, even though it is clear that he might.

Criticizability

Third, many philosophers believe that we can be *criticizable* for our character, i.e., for who we are and what we would do if given the chance. How is this different than blameworthiness? It involves similar reactive attitudes, but without the implication that we have done anything wrong.⁸²

⁸² Christopher Kutz discusses what I am calling 'criticizability' when he argues that we can be "accountable" for our character, independently of whether we act wrongly or cause any harm:

The central distinguishing feature of [accountability warranted by] reasons of character is that the relation between agent and harm need not be mediated by either causality or intentional conduct. Harms may be symbolic, standing for elements of character in agents other than those who brought them about. Thus, reasons of character can allow for associative forms of accountability that reasons of conduct and consequence cannot. Indeed, only reasons of character can make sense of the powerful phenomena of purely vicarious guilt and shame. (Kutz, 2000, *Complicity: Ethics and Law for a Collective Age*, Cambridge: Cambridge University Press), p. 43.

Consider, for example, a case of moral luck. You and I are identical twins, as well as best friends. We have the same beliefs, desires, aims, habits, and so on, and, as a result, we are likely to say or do the same things in the same situations. One night, we feel like a drink. We go to the same bar, have the same number of drinks, and drive the same distance home. You get home safely, but I run over a young child. We can grant for the sake of argument that you are not directly or indirectly *blameworthy* for my running over a child. Nevertheless, you would have done the same thing in the same situation; it was just a matter of luck that the child jumped out in front of my car and not in front of yours. Thus, we think, there is a sense in which you and I are equally *criticizable* here.

Of course, in this case we can at least point to a wrong action that both of us performed: we both drove drunk – an action whose foreseeable consequence is that we might run over a child. But actually, the present point runs deeper than this. Consider another example. You and I feel like a drink. We would both go to the same bar, have the same number of drinks, and drive the same distance home if we could. However, you never pass a bar on the way home. As a result, I drink and drive and run over a child, whereas you make it home safely. In this case, you never act

We can also read Peter Strawson as discussing what I am calling ‘criticizability’ in his famous lecture “Freedom and Resentment.” Strawson argues that reactive attitudes like blame and resentment are directed not just at the *actions* that people perform, but also at the *attitudes* that they express in performing these actions. For instance, if I blame you for shooting me, then I am reacting not just to your shooting me, but also to your being *disposed* to shoot me, as evidenced by your doing so. See Strawson, “Freedom and Resentment,” *Proceedings of the British Academy*, Vol. 48, 1960.

wrongly at all. Nevertheless, you *would* have acted wrongly (and run over a child) if given the chance. Thus, we think, there is still a sense in which we are equally criticizable.

Finally: this point holds to a degree even if you and I are different in some respects, and these differences are enough to make it the case that I would decide, all things considered, to act wrongly, whereas you would decide, all things considered, not to. For example, I might be just a bit more impulsive than you, and, as a result, there might be situations where I would choose to drink and drive and you would not. This might make me a bit more criticizable than you: the more impulsive one is, the more criticizable one deserves for this. Nevertheless, there are still plenty of situations where you would make this decision too, and therefore you are still criticizable to a degree.

Now, suppose that your family knows all this about us: they know that we have many of the same character traits, and, as a result, they know that if I do something, then this is very strong *evidence* that you would do the same thing in the same situation, or at least be tempted to. If so, then there is a sense in which you can be criticizable *in light of* what I do. For example, we can imagine your family exclaiming, “What does it say about you that your *identical twin* would drink and drive and kill a child?” In this case, your family would not be blaming you for my accident. Rather, they would be criticizing you for your own disposition to drink and

drive and run over a child if given the chance – as evidenced by my doing so, and our having many of the same dispositions in general.

How does this apply to personalities? In the same way. Suppose that your “friend personality” drinks and drives, and then your “family personality” comes home and tells your family what happened. Now suppose that your family knows that your personalities have many of the same character traits, and, as a result, they know that if your friend personality does something, then this is very strong *evidence* that your family personality would do the same thing in the same situation, or at least be tempted to. If so, then there is a sense in which your family personality is criticizable *in light of* what your friend personality does. For example, we can imagine your family exclaiming, “What does it say about you that, when you get together with your friends, you turn into someone who would *drink and drive and kill a child?*” In this case, your family would not be blaming your family personality for your accident. Rather, they would be criticizing your family personality for his own disposition to drink and drive and run over a child if given the chance – as evidenced by your friend personality’s doing so, and your personalities’ having many of the same dispositions in general. (They might also, of course, be accusing your family personality of indirect responsibility for what your friend personality did. This is an important point to keep in mind: in typical cases, one personality will be ‘morally responsible’ for what another does in *all*, maybe even *most*, of the senses that I am considering here at once. In my view, this is part of what makes it so hard for us to

keep these senses of ‘moral responsibility’ apart in practice, even in cases where only some of them apply.)

Liability

Fourth, many philosophers believe that we can be *liable* for what other agents do. This means that we can have an obligation to *take* responsibility for what they do, and others can be justified in *holding* us responsible for what they do, even if we did not perform the relevant action, even if we could not have foreseen it, and even if we would not have done the same thing in the same situation (and would not even have been tempted to). The concept ‘liability’ has a legal as well as a moral sense. For present purposes, I will restrict my focus to the moral sense of ‘liability.’ Specifically, I will discuss three ways in which we can be morally liable for what others do.

Commitments

First, we can *commit* to being liable for what another agent does. For example, if someone is thinking of loaning me money, you might *vouch* for me, where this involves making a promise to pay them back if I fail to do so myself.

One of the main ways in which we commit to taking responsibility for what others do is by agreeing to serve as *representative* of a group that we are all members

of. When we play this role, we agree to take responsibility, on behalf of the group, for what other members of the group say or do. For example, if you agree to become President of the United States, then this involves making a commitment to take responsibility, on behalf of your country, for what other Presidents have said and done, on behalf of your country. Similarly, if you agree to become a parent, then this involves making a commitment to take responsibility, on behalf of your children, for what they say or do. Thus, for example, the President might find that he has a moral obligation to apologize, on behalf of his country, for a war that he was not directly or indirectly responsible for prosecuting (and would not have even been tempted to prosecute, if in the same situation). Similarly, a parent might find that he has a moral obligation to apologize, on behalf of his children, for a broken window that he was not directly or indirectly responsible for throwing a rock through (and would not have even been tempted to throw a rock through, if in the same situation).⁸³

How does this apply to personalities? In the same way. Suppose that you have multiple personalities, and each personality makes a commitment to speak and act on behalf of the group – i.e., on behalf of *you*, as a person – in interactions with other people. If they do this, then each is committed to taking responsibility, on behalf of the person, for what other personalities say or do. Thus, for example, your family

⁸³ There are important questions about whether we make these commitments just by performing certain actions. For example, if two people have sex, do they thereby make a commitment to take responsibility for what their child does, in the even that their act results in their having a child? This is a difficult question; fortunately we can set it aside for the time being.

personality might find that he has a moral obligation to apologize, on behalf of the person, for an affair that he was not directly or indirectly responsible for having (and would not have even been tempted to have, if in the same situation).

Fairness

Second, we might have a *fairness-based* obligation to take responsibility for what another agent does, and others might have fairness-based justification to hold us responsible for what another agent does, whether or not we are directly or indirectly responsible for this action, whether or not we would have done the same thing in the same situation (or even been tempted to), and whether or not we made a commitment to take responsibility for this action. Here are just two examples of how.

First, if you benefit from a wrong/harmful action, then you have a fairness-based obligation to compensate those who were wronged/harmed by it. For example, Henry Shue argues that people currently living in industrialized nations benefit from the effects of industrialization, and therefore they have a duty to compensate others who were, or will be, harmed by either industrialization or its effects.⁸⁴ Similarly, suppose that you inherit a large sum of money, and then you discover, after the fact, that your parents earned this money in the drug trade. Many people would claim that

⁸⁴ Shue, 1999, "Global Environment and International Inequality," *International Affairs*, 75: 3, pp. 531-545.

you have a duty to return the money or, if this is impossible, to donate it to a charitable cause, ideally one that benefits victims of the drug trade.

Second, if you *have* more, then you have an obligation to *give* more. Thus, if you benefit materially (or otherwise) from an action, then you might have an obligation to give more as a result of this action, whether or not this action was wrong/harmful. For example, Shue argues that industrialized nations are much *wealthier* than developing nations, and, therefore, they have an obligation to contribute more to the global effort to curb climate change (quite independently of the fact that they got this money as a result of industrialization). Similarly, suppose that you inherit a large sum of money that your parents earned in a perfectly legitimate way. In this case, we might think that you still have an obligation to donate more than me, not because you or your parents did anything wrong, but rather just because you have more money than me as a result of their actions.⁸⁵

How does this apply to personalities? In the same way. First, suppose that your “retirement personality” inherits a large sum of money that your work personality earned in the drug trade. In this case, your retirement personality has a fairness-based obligation to return the money or, if this is impossible, to donate it to a charitable cause, ideally one that benefits victims of the drug trade. Similarly, suppose that your retirement personality inherits a large sum of money that your

⁸⁵ Of course, there are open questions. Most importantly: how much do we have to give in order to discharge these fairness-based obligations? But what matters for present purposes is just that we have to give *something*; and the worse the crime that we benefited from and/or the more resources we have, the more we should give.

work personality earned in a perfectly legitimate way. In this case, your retirement personality still has a fairness-based obligation to donate more than my retirement personality, not because any of your personalities did anything wrong, but rather just because your retirement has more money than mine as a result of your work personality's actions.

Punishing or Harming the Innocent As a Byproduct of Punishing the Guilty

Finally (this is not a sense in which we use the term 'liability,' but it falls under the general heading of 'liability' as defined above), we can have an obligation to take responsibility for what other agents do, and others can be justified in holding us responsible for what other agents do, whether or not we are directly or indirectly responsible for the relevant action, would have done the same thing in the same situation, made a commitment to take responsibility for it, or benefited from it at all.

Of course, if we accept a consequentialist moral theory (or even a deontological theory that has a "threshold" above which consequentialist reasoning is permissible or required), then we can sometimes justify punishing the innocent as a *means* to punishing the guilty, or even as a means to promoting the common good –

especially if the stakes are high enough.⁸⁶ I think this is plausible, both at the interpersonal level and especially at the intrapersonal level. Nevertheless, I will restrict my focus here to a less controversial justification for punishing, or otherwise harming, the innocent, namely: we are sometimes justified in blaming and punishing, or otherwise harming, the innocent as a *byproduct* of punishing the guilty. This is true in an epistemic as well as in a practical sense.

First, we are sometimes justified in blaming and punishing the innocent as a byproduct of blaming and punishing the guilty on epistemic grounds. In particular, we often have to decide whether to blame and punish somebody for a crime without knowing for sure if they committed it. This of course a difficult decision, because if we reserve blame and punishment for people we *know for sure* are guilty, then we will inevitably let many guilty people go free as a result; yet if we allow ourselves to blame and punish people for crimes without knowing for sure that they committed them, then we will inevitably blame and punish at least some innocent people as a result. So the question is: how many guilty people are we willing to let go free for each innocent person we blame and punish? And the answer is usually: a lot, but not an infinite amount. (After all, we need to blame and punish at least *some* guilty

⁸⁶ For example, some have argued that we are justified in hanging an innocent man in order to keep the townspeople from rioting and killing many more innocent people. Similarly (but at the intrapersonal level), Derek Parfit has argued that consequentialist reasoning might justify punishing our future selves for what our past selves did, since many of us have self-interested concern for our future selves, and we might therefore be dissuaded from acting wrongly if we believe that our future selves will be punished for our actions. See Parfit, 1986, *Reasons and Persons* (Oxford: Oxford University Press), pp. 323-6.

people.) Consequently, we settle on a policy for blaming and punishing the guilty whose foreseeable consequence is that we will accidentally punish at least *some* innocent people as well. Of course, it is tragic whenever this happens. But that does not mean that we are wrong to settle on this policy, or to blame and punish people as the policy dictates. Rather, it just means that we live in a complicated world, we have to blame and punish at least some guilty people, and we often have to make decisions about blame and punishment without full information about the case.

Second, even if we *know* who is guilty and who is innocent, we might still have to punish, or otherwise harm, the innocent as a byproduct of blaming and punishing the guilty for *practical* reasons. For example, suppose that a man commits a murder. Now suppose that this man has a conjoined twin (who had nothing to do with his crime). In this case, we know which party is guilty and which is innocent. Nevertheless, we have a very difficult decision to make: either we put the guilty party in jail and also, thereby, put the innocent party in jail, or we let the innocent party go free and also, thereby, let the guilty party go free. In many cases, we might decide that the crime is mild enough that we should just let them both go free. But in at least some cases, we might decide that the crime is serious enough that we should put them both in jail. And even though conjoined twins are somewhat rare, there are many other, less stark examples of punishing or otherwise harming the innocent as a byproduct of punishing the guilty in everyday life. For example, suppose that a man is a good father but a bad husband, and his wife has to decide whether to leave him,

thereby punishing her husband *as well as* harming her children, or stay with him, thereby sparing her children from harm *as well as* sparing her husband from punishment. Either way, as before, this is a tragic choice. But it is also a choice that many of us have to make, since, again, we live in a complicated world, and we will not always be able to make decisions about blame and punishment in a social vacuum.

This latter point is especially important in light of the fact that, in a world with eight billion people and counting, we are not always able to interact with people as *people*. Instead, if we want to have any hope of being able to live and let live on a grand scale, we have to gather into groups of increasing generality –into families, communities, states, nations, and so on – and then interact with groups as *groups*. This has at least two important upshots. First, we might sometimes have a moral obligation to speak and act on behalf of our group in interactions with other groups (whether or not we made a commitment to). For example, if a president promises that his nation will reduce carbon emissions, then his successors have a duty to keep this promise, not only because they made a commitment to do this, but also because they *have* to do this in order to uphold the convention whereby nations interact as nations, as well as in order to meet the expectations that this convention gives rise to. Similarly, if a president wages an unjust war, then his successors might have an obligation to apologize, on behalf of the nation, for this war, not only because they made a commitment to do this (if indeed they did), but also because they *have* to do

this in order to uphold the convention whereby nations interact as nations, as well as to meet the expectations that this convention gives rise to.

Second, and correspondingly, we might sometimes be morally justified in treating other groups as groups, rather than as collections of people, for certain purposes. For example, America might have to impose economic sanctions on China, thereby harming all its citizens, not because they all did anything wrong, but rather because *their government* did; and, for better or worse, we have to treat nations as nations when it comes to economic sanctions. Similarly, we might have to exclude your family from a wedding, thereby harming all your family members, not because they all did anything wrong, but rather because *your son* did; and, for better or worse, families have to be treated as *families* for purposes of wedding invitations. Thus, we will often have to punish or harm some people as a result of the actions or character of others, on the grounds that they happen to be members of the same group (sometimes by choice, sometimes not). Our justification for this is that we have to do it in order to treat groups as groups. And ultimately, our justification for *this* is that it allows us to live and let live easily and efficiently, on a grand scale (an aspiration that we can justify from the standpoints of justice and utility alike).

How does all this apply to personalities? Take each consideration in turn. First, we are sometimes justified in blaming and punishing the innocent as a byproduct of blaming and punishing the guilty on *epistemic* grounds. In particular, we might have to decide whether to blame and punish a personality for a crime without

knowing for sure that they committed it. This is a difficult decision for the same reason as before: If we reserve blame and punishment for personalities who we *know for sure* are guilty, then we will inevitably spare many guilty personalities. Yet if we allow ourselves to blame and punish personalities for crimes without knowing for sure that they committed them, then we will inevitably blame and punish at least some innocent personalities as a result. So the question is: how many guilty personalities are we willing to spare for each innocent personality we blame and punish? As before, we might decide: a lot, but not an infinite amount. Consequently, we might settle on a policy for blaming and punishing guilty personalities whose foreseeable consequence is that we will blame and punish at least some innocent personalities as well. As in the intrapersonal case, it will be tragic whenever this happens. But this will not mean that we are morally wrong to settle on this policy, or to blame and punish personalities as the policy dictates. Rather, it just means that we live in a complicated world, we have to blame and punish at least some guilty personalities, and we often have to make decisions about blame and punishment without full information about the case.

Second, even if we know exactly which personalities are guilty and which are innocent, we might still have to punish, or otherwise harm, innocent personalities as a byproduct of blaming and punishing guilty personalities, for practical reasons. For example, suppose that your “nightclub personality” commits a murder. We might know which personality is guilty and which are innocent. Nevertheless, we have a

very difficult decision to make: either we put the guilty personality in jail and also, thereby, put the innocent personalities in jail, or we let the innocent personalities go free and also, thereby, let the guilty personality go free. In some cases, we might decide that the crime is mild enough that we should just let the whole person go free. But in other cases (like this one, presumably), we might decide that the crime is serious enough that we should put the whole person in jail. Similarly, suppose that a man has a good “father” personality but a bad “husband” personality, and his wife has to decide whether to leave him, thereby punishing or harming both personalities (as well as her children), or stay with him, thereby sparing both personalities (as well as her children). Either way, as before, this is a tragic choice. But it is also a choice that many of us have to make, since, again, we live in a complicated world, and we will not always be able to make decisions about blame and punishment in a social or psychological vacuum.

As in the interpersonal case, this latter point is especially important, since, in a world where many people have multiple personalities in a non-pathological sense, we will not always be able to interact with their personalities as *personalities*. Instead, if we want to have any hope of being able to live and let live on a grand scale, our personalities have to gather into groups, i.e., into *people*, and then interact with people as *people*. This has the same two upshots as before. First, our personalities might sometimes have a moral obligation to speak and act on behalf of the person as a whole in interactions with other people (whether or not they made a commitment

to). For example, if your “family personality” promises to pick your kids up from school, then your “work personality” has a duty to keep this promise (or at least help your friend personality keep it), not only because your work personality made a commitment to do this (if indeed he did), but also because he *has* to do this in order to uphold the convention whereby people interact as people, as well as in order to meet the expectations that this convention gives rise to. Similarly, if your “friend personality” punches someone in the face, then your “family personality” might have an obligation to apologize, on behalf of the person, for this action, not only because he made a commitment to do this (if indeed he did), but also because he *has* to in order to uphold the convention whereby people interact as people, as well as to meet the expectations that this convention gives rise to.

Second, and correspondingly, we might sometimes be morally justified in treating other people as people, rather than as collections of personalities, for certain purposes. For example, I might have to exclude you from my wedding, thereby harming all your personalities, not because they all did anything wrong, but rather because your “nightclub personality” did; and, for better or worse, people have to be treated as people (or maybe even, as I suggested a moment ago, as members of families), for purposes of wedding invitations. Thus, we will often have to punish or harm some personalities as a result of the actions or character of other personalities, on the grounds that they happen to be members of the same person. Our justification for this is that we have to do it in order to treat people as people. And ultimately, our

justification for *this* is that it allows us to live and let live easily and efficiently, on a grand scale.

Importantly, if you are liable for an action for any of the three reasons I have discussed, that does not imply that you are *blameworthy* or even *criticizable* for anything. For instance, we may be justified in calling on America to contribute more in the global effort to curb climate change, or in calling on a President to apologize, on behalf of his nation, for an unjust war that his predecessors waged. But as long as they do this, they will not be blameworthy or criticizable at all, at any step in the process. Similarly, we may be justified in calling on your “retirement personality” to donate money to victims of the drug trade, and to apologize, on behalf of the person, for what your “work personality” did. But as long as he does this (and is otherwise blameless for what your other personalities did), he will not be blameworthy or criticizable at all, at any step in the process.

3.2. Personal Responsibility

I will now argue that, in light of all the different senses in which an agent can be ‘morally responsible’ for an action, my conception of agency can accommodate our judgments about personal responsibility at least as well as other, traditional

conceptions of agency. It implies that, if you have multiple personalities in a non-pathological sense, then, in typical cases, your personalities are “fully” morally responsible for what each other does. And in other, less typical cases where it implies that they are only “somewhat” morally responsible for what each other does, this is, on reflection, the right result.

Take each of these points in turn.

First, my conception of moral agency implies that, if you have multiple personalities in a non-pathological sense, then, in typical cases, your personalities are fully responsible for what each other does (where by ‘fully morally responsible,’ I mean blameworthy, criticizable, *and* liable). After all, if you have multiple personalities in a non-pathological sense, then your personalities share a body/brain, a chain of memory, a very high degree of psychological connectedness, and a self-conception. And each of these relations makes them much more likely to (a) share intentions, (b) be able to foresee how each other will act, (c) have many of the same character traits, (d) commit to being liable for what each other does, (e) benefit from what each other does, and (f) have an obligation to speak and act on behalf of the person as a whole, in interactions with other people. To see why, briefly consider each relation in turn.

3.2.1. First, if your personalities share a body and brain, then they live in the same *territory*. Thus, they have reason, and a moral obligation, to organize into a group,

i.e., a person, and jointly commit to a way of life that will allow them to (a) get along with each other and (b) get along, as a person, with other people. This has implications for all of the senses of ‘moral responsibility’ we have considered. First, they will be more likely to share intentions, since, if they jointly commit to a shared way of life, then each will intend that they live this way together.⁸⁷ Second, they will be more likely to be able to foresee what each other will do, since they will interact more often, and will therefore know more about how each other is likely to act. Third, they will be more likely to have many of the same character traits, since their shared way of life will bring them closer together. (Of course, having the same body, with all its strengths and limitations, will bring them very close together as well.) Fourth, they will be more likely to commit to taking responsibility for what each other does, since, just as agreeing to taking responsibility on behalf of your nation is a precondition for becoming President, agreeing to take responsibility on behalf of the person is a precondition for engaging in many of the projects and relationships that make life worthwhile. Fifth, they will be more likely to benefit (or suffer, as the case may be) as a result of what each other does, since, after all, they will share territory, resources, projects, relationships, and, generally speaking, a common fate. Sixth, other people will often be justified in punishing some personalities as a

⁸⁷ Of course, as with any group, the boundaries of your group will usually be different than the boundaries of the shared agent that corresponds to that group, since there will usually be agents who find themselves in the group yet share none of the relevant intentions, and there will also usually be agents who share some of the relevant intentions yet find themselves excluded from the group. But generally speaking, each group will correspond to a shared agent.

byproduct of punishing others. After all, they will have the same body and voice, which makes it hard to tell them apart, as well as to punish one without also, thereby, punishing the others. And for all these reasons and more, your personalities will have reason, and a moral obligation, to speak and act as a *person*, rather than as a collection of personalities, whether or not they made a commitment to do so (which, again, they probably have).

3.2.2. If your personalities share a chain of memory, then each will remember and anticipate everything that the others say and do (except during blackouts, etc.). This has implications for all the senses of ‘moral responsibility’ we have considered too. First, your personalities will be able to share intentions/actions much more *easily*, since, instead of *communicating* their intentions to each other, they can just form intentions and expect each other to *remember* them. Second, they will be better able to foresee what each other will do, since they will remember everything that each other has done as well as everything that each other is *planning* to do. Third, they will be more likely to have the same character traits, because they will share many more memories, as well as many more beliefs, desires, and other psychological states as a result. Fourth, they will be more likely to commit to taking responsibility for what each other does, because they will be much more likely to think and act like a team in general. Fifth, they will be more likely to be benefit (or suffer, as the case may be) as a result of what each other does, because they will be able to experientially remember

and anticipate each other's pleasures and pains. Sixth, it will be harder for other people to tell them apart, since the transition from one to the next will be more fluid; and it will also be harder for other people to punish some without also, thereby, punishing or harming others, since, again, they can experientially remember and anticipate each other's pains. And for all these reasons and more, they will have even stronger reason, and an even stronger obligation duty, to speak and act as a person in interactions with other people.

3.2.3. If your personalities share a very high degree of psychological connectedness, then they will have many of the same beliefs, desires, habits, and other psychological states as each other. This has many of the same effects as sharing a chain of memory. First, your personalities will be able to share intentions more easily, since they will agree more about what to do and, therefore, can spend less time resolving disagreements and more time working together to promote shared aims. Second, they will be more likely to be able to foresee what each other will do, since they will relate to each other more, and will therefore understand each other better. Third, and obviously, they will have more of the same character traits. After all, sharing a high degree of psychological connectedness *just is* having many of the same character traits (plus having a relation of causal dependence between them). Fourth, they will be more likely to commit to taking responsibility for what each other does, since, once again, they will be more likely to think and act like a team in general. Fifth,

they will be more likely to benefit as a result of what each other does, since, insofar as they agree more about what to do, they will *enjoy* the memory of each other's pleasures more. Finally, it will be even harder for other people to tell them apart and to punish some without also, thereby, punishing or harming others, for the same reason. And for all these reasons and more, they will have even stronger reason, and an even stronger obligation, to speak and act as a person in interactions with other people.

3.4.4. Lastly, if your personalities share a self-conception, then they have a conception of themselves as a *person*, and they each use the first-person to refer to themselves as a person. (They might sometimes think of themselves as personalities as well, but this will be the exception to the rule, and they will usually use terms like 'my family self,' 'my work self,' and so on to refer to themselves as personalities.) This has several important implications too. First, it makes it *extremely easy* for your personalities to share intentions. Why? First, as many philosophers and psychologists have observed, we decide what to do on the basis of our self-conceptions, i.e., we perform the actions that we think it makes sense for us to perform, based on our conception of who we are and what we care about. Thus, if your personalities share a self-conception, then they will tend to form the same intentions. And if they form these intentions from the standpoint of themselves as a *person* (i.e., a *group* of personalities), then they will *share* these intentions naturally,

without even thinking about it. Second, they will be more likely to be able to foresee what each other will do, for the same reason. Third, they will have more of the same character traits, since they will share a point of view and, as a result, will eventually become more alike. Fourth, they will be more likely to commit to taking responsibility for what each other does, since they will think of themselves as committing to taking responsibility for what *they* do. Fifth, they will be more likely to benefit (or suffer, as the case may be) as a result of what each other does, since they will not only experientially remember and anticipate each other's pleasures and pains, but will also think of themselves as the *subject* of these pleasures and pains. Finally, it will be even harder for other people to tell them apart and to punish some without also, thereby, punishing or harming others, for the same reason. And for all these reasons and more, they will have even stronger reason, and an even stronger obligation, to speak and act as a person in interactions with other people.

The upshot is that if you have multiple personalities in a non-pathological sense, then it is *very likely* that your personalities will be morally responsible for everything that each other does, in all or at least most of the senses of 'moral responsibility' that we have considered, i.e., blameworthy, criticizable, and liable. Thus, if my conception of agency is correct, then it explains why other, stronger conceptions of agency *seem* plausible: they seem plausible because each identifies a relation that makes agents very likely to share agency as well as moral responsibility.

To see how this works out in practice, consider two typical cases of adultery involving a culprit with multiple personalities in a non-pathological sense.

1. A man loves his family, and he thinks of himself as fully committed to them. But he also feels a bit stifled by them, and, as a result, his commitment to them is a bit weaker than he thinks – not so weak that he consciously plans to have an affair, but weak enough so that he *unconsciously* plans to have one, i.e., he does everything that he does at least in part because it might lead to an affair (though he may not realize this). For example, he eats well, dresses well, goes to the gym, and so on at least in part so that he can attract potential partners; he goes to the bar with his friends after work at least in part because he might meet someone there, and so on. Then, one night, it all pays off: he meets someone, they start talking, one thing leads to another, and he has an affair. Then he destroys the evidence, goes home, and lies to his family about where he was and what he did. Finally, he settles back into his normal routine. He never has another affair, but this is just by chance. He often thinks fondly about the one that he had, and he fully intends to have another one if given the chance (though, again, he might not realize this).

2. A man loves his family, and he is *usually* fully committed to them. But when he goes to the bar with his friends, a different side of him comes out. In this context, he still loves his family, and he still *thinks* of himself as fully committed to them, but he also feels a bit stifled by them, and, as a result, his commitment is a bit weaker

than normal – not so weak that he consciously plans to have an affair, but weak enough so that he *unconsciously* plans to have one, i.e., he does everything that he does, at the bar, at least in part because it might lead to an affair (though he may not realize this). Most of the time, of course, nothing happens: he goes home, regains his normal perspective, and feels grateful to be back with his family again. But then, one night, he goes to the bar and meets someone new, they start talking, one thing leads to another, and he has an affair. Then he “comes to his senses,” feels terrible about what happened, destroys the evidence, and goes home and lies to his wife – not because he wants to keep having this or any other affair, but because he loves his family and worries about losing them if they find out. Finally, he settles back to his normal routine. He never has another affair, but this is just by chance. As before, he is *usually* fully committed to his family, but he still, on occasion, places himself in contexts where this other personality comes out – and in these moments, if he met the right person, then he would have another affair.

In both these cases, my conception of agency implies, along with other conceptions, that the man is fully morally responsible for the affair, even when he comes back home. In the first case, all of his personalities share an intention to have an affair (whether or not they realize it) and, therefore, they are collectively blameworthy for it, i.e., the man is blameworthy as a *person* for it. Moreover, they all acted in ways that foreseeably caused/allowed the affair to happen, and therefore they are all

indirectly responsible for it as well. (They might not have *foreseen* the affair, but they should have.) Further, they all had many of the same character traits and *would* have done the same things in the same situations; and therefore they are all *criticizable* in light of the affair. Finally, they are all *liable* for the affair, in all the ways we considered: they are all committed to taking responsibility for what each other does⁸⁸, they all benefited from the affair (either by sleeping with another person or by first-personally anticipating or remembering it). And, in light of this, they all have an obligation to *take* responsibility for the affair, and other people are justified in *holding* them responsible for the affair – in part because other people might not be sure which personalities are guilty, and in part because it would be hard for other people to punish some personalities without also, thereby, punishing or harming the others.

Thus, my conception of agency does not imply that, in this case (and others like it), the man can permissibly defend himself, when he gets home from having an affair, by saying that another personality did it. It may be true that another personality slept with someone else, but he is still blameworthy as a *person* (i.e., as a collective agent), as well as blameworthy, criticizable, and liable as a personality.

⁸⁸ This might not always be true, but it will be for most people with multiple personalities in a non-pathological sense. After all, if you have multiple personalities in a non-pathological sense, then your personalities will usually identify with the person as a whole, rather than with themselves as personalities. Thus, they might not *think* of themselves as being committed to taking responsibility for what other personalities do. But in agreeing to take responsibility for what they do as a person, they take on this commitment anyway.

The second case is similar, but has to be treated a bit differently. In this case, we have to distinguish the personality who had an affair from all the other personalities. The “cheating personality” is responsible for the affair in all the ways that I just mentioned: he is blameworthy, criticizable, and liable for it. In contrast, the “normal personalities” did not share an intention to have an affair, and therefore they do not share blameworthiness for it, i.e., the man is not blameworthy as a *person* for the affair. However, many of the normal personalities did share an intention to *cover up* the affair, and they are *indirectly* responsible for the affair as well as the cover up. Moreover, they have many of the same character traits as the cheating personality, and would have been at least *tempted* to do the same thing in the same situation; and therefore they are at least *somewhat* criticizable in light of the affair. (Though we should keep in mind, when deciding how criticizable they are, that very many of us would probably at least be *tempted* to have an affair if given the chance.) Finally, they are all liable for the affair in all of the ways that I just mentioned, except for one: unlike before, they feel terrible about the affair and, as a result, benefit less from the memory of it.

Thus, once again, my conception of agency does not imply that the man can permissibly defend himself, when he gets home from having an affair, by saying that another personality did it and leaving at that. It may be true that another personality had an affair, but his current personality is still blameworthy, criticizable (to a degree), and liable.

In typical cases, then, my conception of agency is fully capable of accommodating our judgment that we are morally responsible for what we did in the past, even if a different personality was out when we performed this action. Granted, it might not, as other conceptions do, always make us *blameworthy*, in each context, for what we do in others. But given that it makes us *collectively* blameworthy for what we do in general and/or *indirectly* blameworthy for what we did in other contexts, *criticizable* in light of what we do in other contexts, and *liable* for what we do in other contexts, who cares?

My conception of agency will not always have this implication, however. In at least some cases, a personality will, despite all the pressure to share intentions with his fellow personalities that I just mentioned, perform an action that his fellow personalities never intended to perform, never could have foreseen, would not have performed themselves, never benefited from, and so on. And in these cases, my conception of agency implies that these other personalities are only *somewhat* morally responsible for this action (where by 'somewhat morally responsible,' I mean somewhat criticizable and liable, and not blameworthy). But I believe that, on reflection, this is the right result. Consider, for example, another case of marital infidelity.

3. A man loves his family, and he is always fully committed to them. But then, one day, he goes on a trip to Vegas (something that he never does), and finds that this trip brings out a new side of his character. Specifically, while on the trip, he still loves his family, and he still *thinks* of himself as fully committed to them; but he also feels a bit stifled by them, and, as a result, his commitment is a bit weaker than normal – not so weak that he consciously plans to have an affair, but weak enough so that he *unconsciously* plans to have one, i.e., he does everything that he does, in Vegas, at least in part because it might lead to an affair (though he may not realize this). And, sure enough, he meets someone during the trip, they start talking, one thing leads to another, and he has an affair. Then he “comes to his senses,” realizes what he did, and resolves to do the right thing: he goes home, tells his family what happened, apologizes profusely, and promises to do everything that he can to make it up to them, as well as to make sure he never places himself in another situation where this other side of him might come out and have another affair. And this is exactly what he does.

In this case, my conception of moral agency implies, along with other conceptions, that the man is morally responsible for the affair when he comes home. But it also implies, unlike other conceptions, that he is not *blameworthy* for anything at all when he comes back home. Specifically, the “cheating personality” is, as in the first and second cases, morally responsible for the affair in all relevant respects. However, the

“normal personalities” are morally responsible in only some respects. They did not share an intention to have an affair or to cover one up, and therefore they do not share blameworthiness for either of these things. Moreover, they could not have foreseen that the trip to Vegas would lead to an affair (since, unlike in the first two cases, the man was not even aware that this side to his character existed), and therefore they are not *indirectly* responsible for the affair either. However, they do share many character traits with the cheating personality and would have at least been *tempted* to do the same thing in the same situation, and therefore they are still at least somewhat *criticizable* in light of the affair. Moreover, they are still somewhat *liable* for the affair, in the same respects as in the second case: they have an obligation to take responsibility, on behalf of the person, for the affair, and other people are justified in *holding* them responsible for the affair, for epistemic as well as practical reasons.

Thus, even in this case, my conception of agency does not imply that the man can permissibly defend himself, when he gets home, by saying that another personality had the affair *and leaving at that*. It may be true that another personality had an affair, but his current personality is still criticizable and liable for it, and this will have to take priority when he talks to his family about the affair.

Nevertheless, this might seem like a revisionary implication to some. In particular, it might seem revisionary to say that the man’s normal personalities are not at all blameworthy for the affair. But I believe that, on reflection, it is the right

result. Intuitively, there really is a sense in which this man's cheating personality is *more responsible* for the affair than his normal personalities are. Granted, the man should apologize profusely for the affair and promise to do everything he can to make it up to his wife and never let it happen again. But provided that he does this, it seems completely reasonable for him to add, "I was a different person when I did this," where this statement makes a distinction between the personality currently in charge and the personality who had the affair. And of course, the point of saying this is that these personalities are, with respect to the affair, different agents: the personality currently in charge did not plan the affair, could not have foreseen it, would not have done the same thing in the same situation, did not benefit from it, and so on. Thus, while the man might, at present, be responsible for the affair in several ways, he is not blameworthy for it. And, yes, as long as he takes responsibility for the affair in all the ways just discussed, it certainly seems reasonable, even just, for him to also clarify the respects in which he is *not* responsible.

Indeed, this kind of reply to blame and punishment is common. People often say "I was a different person when I did this," "I was someone else when I did this," "I wasn't myself when I did this," "a different side of me did this," and so on, in order to draw a line between the personality currently in charge and the personality who performed the relevant action – not in order to escape all responsibility, but rather in order to clarify the (weak) sense in which they *are* responsible. And of course, at least

part of the reason why they do this is that other people often respond well to it. For instance, someone might respond to this defense by saying that they hate the “side of you” that did this but still love the “real you” and know that the “real you” would never hurt them like this.

Granted, some people might use this kind of defense more than they should, and, as a result, we might be skeptical whenever we hear it. Still, the fact that people use this kind of defense at all, and that we take it at all seriously when other they do, shows that we accept that blameworthiness can be compartmentalized in exactly the way that my conception of agency implies: we accept that a person can have multiple personalities that share a body/brain, a chain of memory, a high degree of psychological connectedness, and a self-conception, and yet *still*, in some cases, not be praise- or blameworthy for what each other does.

This is even more apparent when we turn to other, more obvious cases of psychological fragmentation. Consider, for example, a man who suffers from depression. His “normal self” does everything he can to keep his “depressed self” at bay: he takes medication, sees a therapist, studiously avoids anything that might trigger his depression, and makes sure that everyone in his life realizes that he might, despite all these efforts, succumb to depression for days or weeks or even months. Then, one day, he succumbs to depression and decides to spend all day in bed. Unfortunately, this means that he misses several important meetings at work. Now suppose that his normal self comes back out a few days later, tells his boss what

happened, apologizes profusely, and promises to do everything he can to make it up to the company and never let it happen again. In this case, it seems, this man's boss is justified in being angry with his depressed personality, and maybe even in firing the *man*, since after all, people need to be treated as *people* for purposes of employment. But it also seems that the boss would be wrong to be at all *angry* with this man's normal self at any stage in this process. If anything, it seems, he ought to pity this personality for being stuck in a person with a personality that will make it hard for anyone to want to employ him.

Similarly, consider a man who suffers from alcoholism. His "normal self" does everything he can to keep his "drinking self" at bay: he attends meetings, works with a sponsor, goes through the steps, studiously avoids anything that might trigger a relapse, and makes sure that everyone in his life realizes that he might, despite all these efforts, relapse. Now imagine two scenarios. First, imagine that he does this effectively, and his drinking self never comes back out: in this case, it seems that, far from *blaming* his normal self for what the drinking self did, we should *praise* him for doing what it took to punish and restrain his drinking self so that we could begin interacting with him again. Alternatively, imagine that, one night, he relapses, stays out all night drinking, and decides to skip work the next day. Unfortunately, this means that he misses several important meetings at work. Then, a few days later, his normal self comes back out, tells his boss what happened, apologizes profusely, and promises to do everything he can to make it up to the company and never let it

happen again. In this case, as before, it seems that this man's boss is justified in being angry with his drinking self, and maybe even in firing the *man*, since, after all, people need to be treated as *people* for purposes of employment. But it also seems that he would be wrong to be at all *angry* with this man's normal self at any stage in this process. If anything, as before, it seems that he ought to pity this personality for being stuck in a person with a personality that will make it hard for anyone to want to employ him.

These cases illustrate especially well the tragedy of being trapped in a person with a "bad" personality, as well as the injustice of being blamed for what this personality does when you had nothing to do, directly or indirectly, with the relevant action, would never do the same thing in the same situation, and, indeed, did everything you could to stop them. And of course, the only real difference between these cases and the infidelity cases is that, in these cases, the psychological changes that the man undergoes are more significant, and, therefore, (a) the innocent personalities are a bit less criticizable in light of the relevant action, and (b) other people are a bit less justified in blaming and punishing the innocent personalities on epistemic grounds (since it is a bit easier for them to tell each personality apart).⁸⁹

⁸⁹ Some philosophers believe that there is another difference between these cases: a depressed or drinking personality is not an *agent*, and is therefore not blameworthy for what they do. I think that a very weak version of this claim might be true: it might be that, insofar as a person is depressed or drunk, they might be less than *fully* autonomous. But they are clearly still autonomous to a very high degree. Thus, I think the best way to explain our judgments about these cases is not to say that that

3.3. Conclusion

On my conception of agency, if a person has multiple personalities in a non-pathological sense, then each personality is, at best, praise- or blameworthy only for what he does. This might seem like a revisionary idea. However, I have argued that, even if each personality is praise- or blameworthy only for what he does, a person can still be collectively praise- or blameworthy for what his personalities do together, and each personality can also be indirectly responsible for what the others do, criticizable in light of what the others do, and liable for what the others do. Thus, my conception of agency has plausible implications for responsibility after all.

Finally, let me just repeat that this entire discussion has been based on the assumption that an agent is praise- or blameworthy for what he does. If we reject this assumption, then my conception of agency will have the same implications for personal responsibility as other conceptions: a person is never praise- or blameworthy for anything that they do, though they can still be criticizable and liable for many of their actions. Many philosophers believe that this is a *very* revisionary idea. But one of the upshots of my discussion in this paper is that it might not be as revisionary as we thought. Indeed, if we reduce blameworthiness to causal

the depressed or drinking personality is not an agent, but rather to say that they are a *different* (and perhaps less autonomous) agent than the normal personality.

responsibility, criticizability, and liability, then we can justify many of our judgments about moral responsibility by reference to the latter concepts alone, with the same plausible implications that my conception of agency has here.

Conclusion

I have argued that personalities can be different agents even if they share a body/brain, a chain of memory, a high degree of psychological connectedness, and a self-conception. Thus, for example, if you have a "day self" who likes to work and a "night self" who likes to play, then these personalities count as different agents.

If this is right, then it raises a puzzling question: if none of the relations I just mentioned is sufficient to make your present self part of the same agent as a past or future self, then what relation, if any, is sufficient to do this? More to the point: how is it possible that your present self is part of the same agent as, say, yourself tomorrow morning but not, say, yourself tonight – especially when the psychological difference between “yourself tonight” and “yourself tomorrow morning” is so thin?

It might be tempting to answer this question by accepting a “personality” theory of agency: your present self is part of the same moral agent as a past or future self iff it is part of the same *personality* as them. In my view, however, this view encounters many of the same problems as the “person theory of agency,” i.e., the theory that your present self is part of the same moral agent as a past or future self iff it is part of the same *person* as them. Specifically, it raises tricky metaphysical questions about the nature and persistence conditions of personalities (do they exist? is “personality identity” transitive? all-or-nothing? etc.) – questions which seem not

only difficult to answer, but also ultimately irrelevant to the moral questions we have been considering.

For instance, what if we decide that, as a matter of metaphysical fact, personality identity is transitive, and therefore personalities do not survive fission and fusion? Unlike in the personal identity case, this will have many real-world implications, because personalities undergo fission and fusion all the time. For example, compare the following two cases, which are exactly the same except that one involves fission. In the first, you get a job at a liberal arts college, and as a result, your “teaching” personality gradually changes into an “undergraduate teaching” personality. In the second, you get a job at a research university, and as a result, your “teaching” personality gradually *splits* into an “undergraduate teaching” personality and a “graduate teaching” personality. Practically speaking, your present self might have the same exact relationship with your past selves in both cases. Yet if personality identity is transitive, then your present self might be part of the same personality as your past self in the first case but *not* in the second. But could this plausibly make a moral difference? i.e., is it plausible that, if you get the job at the liberal arts college, then your past self is permitted to bind your present self against his will, and that your present self is blameworthy for his doing so? but that, if you get the job at the research university, this is not the case? It seems clear to me that the answer is no: your past self is wrong to bind your present self against his will,

and your present self is not blameworthy for his doing this, independently of whether, as a matter of metaphysical fact, they are part of the same personality.

If we reject the personality theory on these grounds, then we might be tempted to accept a revised “psychological connectedness” theory of agency instead. This would be similar to the personality theory, but without the implication that morality is governed by the logic of numerical identity (and other such metaphysical issues). On this view, your present self is part of the same agent as a past or future self (independently of whether they are, as a matter of metaphysical fact, part of the same personality) iff they share a *very, very high degree* of psychological connectedness. In my view, however, this theory is problematic too. After all, if you have multiple personalities in a non-pathological sense, then, as I said, the psychological difference between them is very thin. For instance, your present self may share a *few* more beliefs, desires, intentions, habits, and so on with yourself tomorrow morning than with yourself tonight – this is why it helps us to describe your present self as sharing a personality with yourself tomorrow morning but not with yourself tonight – but these differences are likely to be *very minor* compared to the vast number of beliefs, desires, intentions, habits, and so on that your present self shares with both future selves. After all, even if your present self disagrees with yourself tonight about, say, whether to stay out late and have a few extra drinks, they might still agree about, say, whether to brush your teeth twice a day, whether to look both ways before crossing the street, whether to stay in your current relationship, whether to stay at

your current job, and so on. In light of this, it becomes very difficult to say whether and why we should draw a line between yourself tonight and yourself tomorrow morning – why we should say that yourself tomorrow morning shares enough psychological connectedness with your present self to be part of the same agent, but yourself tonight does not.

Of course, in some cases we can just *stipulate* where to draw the line, because all of the examples we are interested in capturing fall clearly above or below our stipulated line, and also because the stakes are not very high to begin with. For example, Derek Parfit has argued that, for purposes of defining personal identity, we can just stipulate that your present self is part of the same person as a past or future self only if they are at least 50 percent psychologically connected, since, (a) in typical cases, your present self will have either a much higher or much lower degree of psychological connectedness with a past or future self, and (b) personal identity is not, for Parfit, what matters in survival, rationality, or morality anyway. But in this case, we would be drawing a line between, say, 99.8 percent psychological connectedness and 99.9 percent psychological connectedness. And any choice along these lines would seem arbitrary, since (a) many of the examples that we are interested in capturing will fall right on or around this line, and (b) the stakes are very high, since where we draw the line will determine whether, say, a past self has a right to bind you against your will, or whether a future self is blameworthy for what you do. Thus, it seems to me that degree of psychological connectedness cannot be,

in itself, what matters either. Intuitively, your past self is wrong to bind your present self against his will, and your present self is not blameworthy for his doing so, independently of whether they happen to meet or exceed our stipulated threshold of psychological connectedness. (Though, as I argued throughout this dissertation, sharing a high degree of psychological connectedness still matters a great deal, since it makes your present self very likely to share intentions with your past selves, to be criticizable in light of what they do, to benefit from what they do, to have an obligation to take responsibility for what they do, and so on.)

If we reject the personality theory *and* the psychological connectedness theory on these grounds, then we might be tempted to accept a revised “self-conception” or “self-narrative” theory of agency instead. On this view, your present self is part of the same agent as a past or future self iff they share an “agent self-conception,” e.g., if they *think* of themselves as part of the same agent. But this theory seems arbitrary too. For instance, what if your present self shares an agent self-conception with the past self who bound him against his will? i.e., what if he describes the situation by saying “I bound myself against my will,” or “my day self bound himself against his will,” rather than by saying, “a different personality bound me against my will,” or “my night self bound my day self against his will”? Would this make it the case that your past self was permitted to bind your present self against his will, or that your present self is blameworthy for his doing so? It seems clear that the answer is no. Your past self is wrong to bind your present self against

his will, and your present self is not blameworthy for his doing so, independently of how they think and talk about it. Indeed, if anything, it seems as though sharing an agent self-conception in this kind of situation can be *harmful*, because it can obscure the moral and rational facts: it might cause your present self to feel *regret* rather than *blame* as a result of what happened, and it might also cause him respond by *intending* not to do this again (an intention that will be thoroughly impotent), rather than by trying to *persuade* or otherwise *compel* your future “day selves” not to do it again. Thus, it seems to me that sharing an agent self-conception cannot be, in itself, what matters any more than sharing a very, very high degree of psychological connectedness can. (Though, as with sharing a very, very high degree of psychological connectedness, sharing a self-conception still matters a great deal, for all the same reasons: it makes your present self very, very likely to share intentions with your past selves, to be criticizable in light of what they do, to benefit from what they do, to have an obligation to take responsibility for what they do, and so on.)

I am not claiming that these considerations are decisive. But they do make me want to look elsewhere for an explanation of how your day self and night self can be separate agents. Fortunately, I believe, there is an alternative theory that avoids all these problems, and also paves the way for a simple, unified, theoretically plausible conception of agency. On this theory, which I call the “temporal self” theory, each of your temporal selves is an individual agent. Each temporal self has the right to live the way that he wants, and is praise- or blameworthy, at best, only for what it does.

However, insofar as your temporal selves share *intentions* with each other, they are also part of a *collective agent*, with the upshot that they (a) have a collective right to perform the actions that they share an intention to perform and (b) are collectively praise- and blameworthy for the actions that they perform together. On this view, *insofar* as your “daytime” temporal selves share intentions with each other, your “day personality” counts as a collective agent; and similarly, *insofar* as your “nighttime” temporal selves share intentions with each other, your “night personality” counts as a collective moral agent. But of course (and this seems like the right result), in cases of non-pathological multiple personality, the truth will usually be much more complicated: your “daytime” temporal selves will sometimes fail to share intentions with each other, your “nighttime” temporal selves will sometimes fail to share intentions with each other, and your “daytime” and “nighttime” temporal selves will very often *succeed* in sharing intentions with each other. Thus, instead of thinking of your “day personality” and “night personality” as completely distinct agents, the “temporal self” view encourages us to think about them as groups of temporal selves who, like members of the same political party, are *especially likely* to share intentions with each other – though they might sometimes fail to share intentions with each other, and they might sometimes succeed in sharing intentions with others.

The temporal self theory of agency, then, explains why your day personality and night personality are (or at least can be) separate agents, while at the same time avoiding all the problems of the other theories I just mentioned. Specifically, it does

not make your obligations to past and future selves depend on (a) whether you are, as a matter of metaphysical fact, part of the same personality, (b) whether you are, say, 99.8 percent psychologically connected to them rather than, say, 99.9 percent psychologically connected to them, or (c) whether you happen to share an agent self-conception with them. Rather, the temporal self theory says that your present self has a duty not to bind *any* future self against its will, and he is not praise- or blameworthy for *any* action that he had nothing to do with. Yet it also implies that, insofar as your present self shares intentions with past and future selves, it has no *reason* to bind them against their will, and it *shares* praise- or blameworthiness with them for what they do together. Thus, the temporal self theory makes the moral facts depend on what, intuitively, they ought to depend on: how your temporal selves treat each other, and what they share an intention to do.

If the temporal self theory is right, then it casts questions about what matters in survival, rationality, and morality in a different light. In particular, it implies that these questions are not, as some philosophers have assumed, about what marks the difference between “self” and “other” in survival, rationality, and morality (since, on the temporal self view, all past and future selves are “other” to our present selves.) Rather, these questions are about what it takes for our present selves to have an especially *intimate relationship* with past and future selves, in a way that makes them more likely to share intentions (as well as other rationally and morally significant relations). And if this is right, then, instead of thinking of standard theories of what

matters in survival, morality and rationality as *competing* answers to the same question – i.e., instead of thinking that *either* sharing a body/brain is what matters, *or* sharing a chain of memory is what matters, *or* sharing a high degree of psychological connectedness is what matters, *or* sharing a self-conception is what matters (where this is an exclusive ‘or’) – we ought to think of them as *complementary* answers to this question – i.e., we ought to think that each of them identifies a relation such that, if our temporal selves share this relation, then they are much more likely to have an especially intimate relationship with each other and, therefore, are much more likely to be “self” to each other *in practice*, even if they continue to be “other” to each other in theory.

I am not claiming to have established that the temporal self theory of agency is true, or that these other theories are false. But I do think that my arguments in this dissertation, as well as my more speculative arguments in this conclusion, point to the temporal self theory as promising. My next step in this project, then, will be to explore the temporal self theory in more detail, along with the more complicated analogy with the state that it gives rise to.

Works Cited

Altrocchi, John, 1999, "Individual Differences in Pluralism in Self-Structure," in John Rowan and Mick Cooper (eds), *The Plural Self*, Sage Publications, London.

Bloom, Paul, 2008, "First-Person Plural." *The Atlantic*.

Bratman, Michael, 1999, *Faces of Intention* (New York: Cambridge University Press, 1999).

———, 2009, "Shared Agency," *Philosophy of the Social Sciences: Philosophical Theory and Scientific Practice*, C. Mantzavinos, ed., New York: Cambridge University Press, 41–59.)

Buss, 2008, "Personal Autonomy," *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), URL = < <http://plato.stanford.edu/entries/personal-autonomy/> >

Carter, Rita, 2008, *Multiplicity: The New Science of Personality, Identity, and the Self*, London: Little, Brown.

Dennett, 1992, "The Self as a Center of Narrative Gravity," in F. Kessel, P. Cole and D. Johnson, eds, *Self and Consciousness: Multiple Perspectives*, Hillsdale: Erlbaum.

Doris, John. 2005, *Lack of Character: Personality and Moral Behavior*, Cambridge: Cambridge University Press.

Dworkin, Gerald, 2005, "Paternalism," *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, URL = <http://plato.stanford.edu/entries/paternalism/>.

Elster, Jon, 2000, *Ulysses Unbound: Studies in Rationality, Precommitment, and Constraints*, Cambridge: Cambridge University Press.

Feinberg, Joel, 1984, *Harm to Others* (New York: Oxford University Press).

Ferrero, Luca, 2010, "Decisions, Diachronic Autonomy, & the Division of Deliberative Labor," *Philosophers' Imprint*, 10: 2.

- Frankfurt, Harry, 1987, "Identity and Wholeheartedness," in Ferdinand David Schoeman (ed.), *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*. Cambridge University Press.
- Gilbert, Margaret, 1996, *Living Together: Rationality, Sociality, and Obligation*, Lanham: Rowman and Littlefield.
- , 2000, *Sociality and Responsibility*, Lanham: Rowman and Littlefield.
- , 1990, "Walking Together: A Paradigmatic Social Phenomenon," *Midwest Studies in Philosophy* 15:1–14
- Goffman, Erving, 1959, *The Presentation of Self in Everyday Life*, New York: Anchor Books.
- Goodin, Robert, 1988, "What is So Special about Our Fellow Countrymen?" *Ethics* 98: 4, pp. 663–686.
- Hacking, Ian, 1998, *Rewriting the Soul: Multiple Personalities and the Science of Memory*, Princeton: Princeton University Press.
- Harman, Gilbert, 2000. "The Nonexistence of Character Traits." *Proceedings of the Aristotelian Society* 100: 223–226.
- Hart and Honore, 1985, *Causation in the Law*, Rev. 2nd ed. (New York: Oxford University Press).
- Herman, Barbara, 1993, *The Practice of Moral Judgment*, Cambridge: Harvard University Press.
- Kant, Immanuel, 1998, *Groundwork of the Metaphysics of Morals*, trans. Mary Gregor, Cambridge: Cambridge University Press.
- , 1996, *The Metaphysics of Morals*, trans. Mary Gregor, Cambridge: Cambridge University Press.
- Kavka, Gregory, 1983, "The Toxin Puzzle," *Analysis*, 43: 1, pp. 33–6.
- Korsgaard, Christine, 2009, *Self-Constitution: Agency, Identity, and Integrity*, Oxford: Oxford University Press.

- , 1996, *The Sources of Normativity*, Cambridge: Cambridge University Press
- Kutz, 2000, *Complicity: Ethics and Law for a Collective Age* (Cambridge: Cambridge University Press)
- Linville, P. W., 1987, “Self-Complexity as a Cognitive Buffer Against Stress-Related Illness and Depression,” *Journal of Personality and Social Psychology*, Vol. 52, pp. 663-76.
- Mill, John Stuart, 2008, *On Liberty and Other Writings*, Cambridge: Cambridge University Press.
- , 2004, *Utilitarianism and Other Essays*, London: Penguin.
- Nietzsche, 1968, *The Will to Power*, trans. Walter Kauffman & R. J. Hollingdale, New York: Vintage.
- Olson, Eric, 1997, *The Human Animal: Personal Identity without Psychology*, Oxford: Oxford University Press.
- Parfit, Derek, 1984, *Reasons and Persons*, Oxford: Oxford University Press.
- Rabin, Nathan, 2011, “Interview with Louis C.K., *The Onion A.V. Club* <URL = <http://www.avclub.com/articles/louis-ck,58516/>>
- Rawls, John, 1999, *A Theory of Justice: Revised Edition*, Cambridge: Harvard University Press.
- Rovane, Carol, 1997, *The Bounds of Agency: An Essay in Revisionary Metaphysics*, Princeton: Princeton University Press.
- Sawyer, R. Kieth, 2001, *Creating Conversations: Improvisation in Everyday Discourse* (Cresskill, NJ: Hampton Press, 2001)
- Schechtman, Marya, 1996, *The Constitution of Selves*, Ithaca: Cornell University Press.
- Shakespeare, William, 2004, *As You Like It* (New York: Washington Square Press)
- Shue, Henry, 1999, “Global Environment and International Inequality,” *International Affairs*, 75: 3, pp. 531-545.

Strasberg, Lee, 1988, *A Dream of Passion: The Development of the Method* (New York: Plume)

Strawson, "Freedom and Resentment" (Proceedings of the British Academy, Vol. 48, 1960).

Street, Sharon, 2009, "In Defense of Future Tuesday Indifference: Ideally Coherent Eccentrics and the Contingency of What Matters," *Philosophical Issues*, vol. 19 on "Metaethics," ed. Ernest Sosa.

Swann, William, 1999, *Resilient Identities: Self, Relationships, and the Construction of Social Identity*, New York: Basic Books.

Taylor, Marjorie and A. Kohanyi, 2002/3, "The Illusion of Independent Agency: Do Adult Fiction Writers Experience Their Characters as Having Minds of Their Own?" *Imagination, Cognition and Personality*, No. 22, pp. 361-80.)

Upton, Candace, upcoming, "The Structure of Character," *Journal of Ethics*.

Velleman, J. David, 2009, *How We Get Along*, Cambridge: Cambridge University Press.

——, 2006, "Self to Self," in *Self to Self: Selected Essays*, Cambridge: Cambridge University Press, pp. 170-202.

——, 1997, "How to Share an Intention", *Philosophy and Phenomenological Research*, 57, pp. 29-50.