

Autonomy and the Authority of Personal Commitments: From Internal Coherence to Social Normativity

Joel Anderson

Abstract

It has been argued – most prominently in Harry Frankfurt’s recent work – that the normative authority of personal commitments derives not from their intrinsic worth but from the way in which one’s will is invested in what one cares about. In this essay, I argue that even if this approach is construed broadly and supplemented in various ways, its *intrasubjective* character leaves it ill-prepared to explain the normative grip of commitments in cases of purported self-betrayal. As an alternative, I sketch a view that focuses on *intersubjective* constraints of intelligibility built into social practices and on the pragmatics of how those norms are contested in an ongoing fashion.

I. Introduction

What it is good for one to do with one’s life clearly depends on who one is. There are several senses in which this is true. Certain careers, relationships, or projects will be fulfilling only if one has the appropriate aptitudes and dispositions. Similarly, certain facts about one’s context – one’s culture, one’s family, one’s race, one’s gender, one’s age, and so on – can represent important considerations to be

taken into account in making significant choices. There is, however, a further way in which who one is determines what one should do, and that is my focus here. In this sense, one’s identity may be centrally constituted by the fact that one cares deeply about a project, person, or ideal, such that betraying this commitment would amount to betraying oneself.

I am thus concerned here with the peculiar sort of claim that some motives, desires, and attachments make on an agent. They give one compelling reasons to disregard many of one’s desires. The fact that one cares deeply about something gives the associated desires authority that other desires lack. What is interesting

about these personal commitments is that the authority involved is highly contingent and personal, in the sense that their claim on one is at least in part a function of what one happens to care about, rather than being exclusively a matter of what deserves, objectively, to be valued. There is no mistake in not wanting to learn to play all the Beethoven piano sonatas, or even in not caring about classical music. There are no general obligations or basic needs or objective interests that require such a commitment. And yet, it is perfectly possible that someone could care about learning to play the Beethoven sonatas such that a failure to pursue that project would represent a serious failure, more serious than not achieving the goal. Because this could involve a failure to lead one's life in accordance with what is most central to one's identity, it could be self-betrayal and thus (because one is failing to guide oneself in light of what really matters to one) a failure of autonomy.

What needs to be explained is why some desires have authority and how they *get* that authority.¹ How is it that desires and personal commitments can get the sort of normative *grip* on one that constitutes authority, such that not acting on them would constitute a shortcoming, a normative failure? How is it that, in some cases at least, personal commitments generate a standard for one, a standard one must heed if one's autonomy is not to be impugned?

In the recent literature on autonomy, nonmoral value, and practical reasoning – especially under the influence of Harry Frankfurt – the authority of desires and personal commitments is seen as depending on their exhibiting a certain degree of *internal* connectedness or coherence. The idea is that disintegration and fragmentation are incompatible with autonomy, since in the absence of such coherence, the agent has no standards or “*nomos*” in accord with which to guide herself autonomously. People whose desires are in constant and unpredictable flux lack the sort of organization required for having agency and having a self. Less familiar is the idea that being autonomous involves being connected in particular ways *to others*. My aim here is to argue that once we properly see how *internal* coherence must be understood if it is to play its important role in theories of autonomy and value, it will become clear that the authority of personal commitments also has a fundamentally *social* dimension.

In the first part of the paper, I reconstruct and criticize an attempt to explain what makes some desires authoritative solely in terms of internal coherence, starting with the “volitional” model found in Harry Frankfurt's recent work, and then considering how the approach could be improved through taking a more “inferentialist” approach to internal coherence.² I focus on the case of self-betrayal, as a way of shedding light on how it is that some of one's desires and personal com-

- 1 Note that in phrasing the question this way, I am leaving to one side broadly naturalist and instrumentalist approaches that start from the assumption that all desires are authoritative until proven otherwise – proven otherwise, for example, by identifying inappropriate causal influences or other external factors. For these theorists, what needs to be explained is how a desire could *fail* to have authority. Brandt (1979) is the classic point of reference; Noggle (1997) is a more recent example.
- 2 E.g., Christman (2001), Ekstrom (1993), Velleman (1989), and the essays by Frankfurt cited below.

mitments can be viewed as having authority over one's choices, such that departing from those commitments can count as a *violation*. The specific shortcomings of the internal coherence approach suggest the approach I develop, an approach focused on norms of intelligibility and norms that govern the social practices in which judgments are made about whether one has betrayed oneself.

2. Harry Frankfurt and caring wholeheartedly

As is familiar, Harry Frankfurt's early essay, "Freedom of the Will and the Concept of the Person" (Frankfurt 1987, ch.2) suggested that autonomous agency is a matter of one's actions flowing from desires that one *reflectively endorses* having. What distinguishes autonomous action and choice, on this view, is the presence of "second-order" attitudes regarding one's first-order desires. As critics quickly pointed out, however, it is unclear what the special authority of second-order desires is: they themselves are subject to all the same autonomy-impairing influences to which first-order desires are subject.³ And attempts to adjudicate the authenticity or "internality" of second-order desires by appeal to *third-order* desires simply generates a regress.

Frankfurt actually noticed this problem early on: "an infinite regress will be generated by any attempt to account for internality or externality in terms of attitudes."⁴ Something more is clearly needed. According to Frankfurt's current approach,⁵ that additional element is a certain degree of integration and equilibrium within the volitional makeup of the person. On this view, our desires, values, commitments, etc. have authority for us in virtue of some form of involvement with and investment in ourselves. In particular, it is the organization of the agent's will that lends clear authority to a particular class of her desire-like aspects, namely, *what she cares about wholeheartedly*. Caring about a project, person, or ideal – or, as Frankfurt also puts it, loving it – is a matter of having a stake in something:

A person who cares about something is, as it were, invested in it. He *identifies* himself with what he cares about in the sense that he makes himself vulnerable to losses and susceptible to benefits depending on whether what he cares about is diminished or enhanced (Frankfurt 1987, p. 83).

Caring *wholeheartedly* is a matter of caring in a way that generates no conflicts in the agent's will: no ambivalence, no hesitation, no tensions. As such, Frankfurt argues, caring wholeheartedly gives one very good reasons to act on the associated desires. These are not the only grounds for autonomous action, but in the

3 Representative here are Watson (1975); Thalberg (1978); and Friedman (1986).

4 Frankfurt (1987), p. 65. (That chapter was originally published in 1976.) Even more clearly: "The mere fact that one desire occupies a higher level than another in the hierarchy seems plainly insufficient to endow it with greater authority or with any constitutive legitimacy" (Frankfurt (1987), p. 166.

5 See especially Frankfurt (1987, ch. 7, 12, 13) and (1999, ch. 7-9, 11, 14).

absence of felt doubts, there is, Frankfurt argues, no reason to deny the authority of these desires.⁶

It is important to note that caring wholeheartedly about a project, person, or ideal is quite different from desiring that a particular state of affairs obtain – even at the level of one’s highest-order desires. According to Frankfurt, in caring about something, one not only wants it to flourish, one also wants it to be the case that one’s own action contribute to – or at least do not detract from – its flourishing.

Insofar as a person loves something, the fact that he cares about it as he does requires that he must care similarly about how he acts in matters that concern it....In the very nature of the case, he cannot be indifferent to how what he does affects his beloved. To the extent that he cares about the object of his love, therefore, he necessarily cares about his own conduct as well (Frankfurt 1999e, p. 138).

In a significant range of cases, then, being volitionally invested in a project, person, or ideal means that one is also volitionally invested in guiding oneself (so as to promote the well-being of what one cares about). This reflexive aspect of what it is to be an agent who cares wholeheartedly provides the basis for saying that one’s caring wholeheartedly about something makes demands on one, such that voluntarily acting against the well-being of an object of care entails the violation of something central to who one is: “A person who fails to act in the ways that caring about his beloved requires necessarily fails to live in accordance with his ideal for himself. In betraying the object of his love, he therefore betrays himself as well” (Frankfurt 1999e, p. 139). For my purposes, what is important about this is that it suggests a way of understanding the “grip” that some desires have, that is, of how one’s choices might be held accountable to a standard that is not reducible to what one happens to desire at a given moment. The fact that one cares deeply and wholeheartedly about something means that some desires are overridden, not as a matter of strength, but of authority. Whatever one ends up doing, one’s desire to ensure the flourishing of what one cares about still has a grip, in that it sets a standard according to which one can count as having betrayed oneself.

Before continuing, it is worth underscoring Frankfurt’s insistence that the authority of one’s commitments comes not from their worth but from how they are connected to one’s *volitional* make-up. Indeed, despite his continuing association with the notion of “second-order desires,” Frankfurt now downplays any requirement of actual reflection. The model of autonomous action now seems to be the form of active and volitional reflexivity that is found paradigmatically in caring about or loving a person, project, or ideal. One does not have to actually

6 This is particularly clear in Frankfurt (1987, ch. 12) and (1999, ch. 8). The appeal to wholeheartedness is important, for it allows Frankfurt to say why the merely psychological “grip” that an obsession (for example) has on one should not be viewed as having authority. Thus, however *forceful* an obsession may be, as long as it is experienced as exerting control over one, its authority will not be that of the sort of commitment compatible with autonomy and free agency (see 1987, p. 183; 1999e, p. 136).

have reflected consciously on one's commitments for them to have authority, but one's will *does* have to be constituted in such a way that one has a stake in one's actions towards what one cares about. Put slightly differently, what makes many of one's actions autonomous is not that the motives underlying them have been endorsed in critical reflection, but rather that one is acting in a way that is subject to "imperatives" of the will, that is, to certain forms of "*volitional necessity*." It is arguably in this sense that Martin Luther said, "Here I stand; I can do no other." As Frankfurt interprets these cases, there are things that we cannot help caring about, not in virtue of thinking them important, but in virtue of how our will is invested in them – and that this does not count against them but in favor of them.⁷

Consider an example Frankfurt discusses (1987, pp. 163-4), taken from Anthony Trollope's novel, *The Eustace Diamonds*, in which an aristocrat, Lord Fawn, has the opportunity to find out from a low-born estate-steward whether his fiancée is cheating on him. He resolves to ask the steward for the information, but then "every feeling in his nature revolts" against discussing such a delicate matter with him. However prejudiced and classist Lord Fawn may be, Frankfurt takes the case to illustrate how all the practical reasoning in the world may still have to answer to fundamental constraints of what one can bring oneself to do – and that being subject to such constraints might be part of what it is to have any identity at all.

This volitional approach has the advantage of avoiding the problems associated with requirements of explicit reflection. On Frankfurt's approach, being autonomous is clearly compatible with a wide variety of lifestyle, including lives of deep devotion.⁸ Because caring is a volitional feature of a person rather than a cognitive achievement, it is possible for a person to count as acting on desires that are authoritative for her even though she has not given them an explicit stamp of approval. On Frankfurt's model, authority stems ultimately from the reflexive involvement of the *will* rather than of the intellect.

This apparent advantage of avoiding a requirement of critical reflection actually becomes a deficit, however, when we ask how self-betrayal is to be identified on volitionalist grounds. If the reason-giving authority of core personal commitments depends entirely on currently caring about the person, ideal, or project in question, and if being committed is of a piece with acting wholeheartedly in support of those commitments, then it seems that acting at odds with a core personal commitment seems to always end up being not a violation but simply evidence that one no longer cares. As soon as one starts caring about something else, the prior commitment has, *de facto*, lost its authority. Thus, if one accepts the volitional approach, it becomes hard to see how commitments could ever have enough traction for it to be possible to betray oneself willingly.

Consider again the case of Lord Fawn, but with a different sequel, in which the

7 On "volitional necessity," see especially Frankfurt (1987, p. 87-88; 1987, ch. 13, 1999, ch. 9 and 11). As Daniel Dennett asks rhetorically, "Doesn't a considerable part of being a responsible person consist in making oneself unable to do the things one would be blamed for doing if one did them?" (Dennett 1984, p. 135).

8 For a further defense of this advantage, see Buss (1994). For a critique, see Friedman (2000).

thought of possibly being cuckolded leads him to approach the estate-steward again. This time, the previous volitional barriers dissolve, and he finds himself surprisingly at ease speaking to the steward. Has he betrayed himself? Frankfurt's own discussion suggests that Lord Fawn's inability to speak to the steward was what kept him from betraying himself. But once the unthinkable has occurred, appeals to volitional necessity turn out to be empty. The authority of Lord Fawn's volitional character that was taken for granted in the Trollope version becomes an open question. There are now three initial possibilities. Lord Fawn might be betraying himself; he might now be acting in line with his authoritative commitments (although he wasn't when he initially refused to speak to the steward), or he might simply have had a change of heart (in which case, he has shown integrity in both cases). The question is what *volitional* criterion there could be for determining which version actually obtains.

One possibility is to argue that ambivalence or internal conflict is a decisive marker for when a person's departure should count as self-betrayal. Frankfurt regularly stresses the ways in which ambivalence stands opposed to the kind of internal integration that is the condition for the possibility of self-guidance.⁹ Thus, if Lord Fawn has been wholeheartedly committed to certain principles of etiquette and a subsequent departure from those principles is accompanied by ambivalence and internal conflict, he would have gone from a healthy state of the mind to a diseased state (Frankfurt 2000a, p. 12), and that might count as self-betrayal.

There are two problems with this approach. First, the charge of self-betrayal here seems to miss its target. For what is betrayed is not the authoritative commitment but rather one's status as an agent. That may be important, but it does not answer the question of which commitment has authority. Second, this approach makes it hard to distinguish self-betrayal from a lack of wholeheartedness, which for Frankfurt is what gives one's caring about something authority in the first place. Thus, if one is simply divided about a person, project, or ideal, there is no reason to think that being divided is a case of violating something that has authority. It may be that one's caring about this person, project, or ideal no longer has the normative hold on one that it once had. At the very least the grip has become uncertain. In short, on this view, it becomes impossible to say that one is betraying something that has an authoritative hold on one, since that requires that one care wholeheartedly about it.

This suggests a second approach, according to which one's departure from a previous commitment counts as self-betrayal only if one experiences shame or regret about it. This regret-based approach is more flexible in that it drops the requirement of feeling *occurrent* wholehearted identification with one's project and relies instead on an affective-volitional rejection, at one point in time, of a course of action that one has taken (wholeheartedly) at another point in time. Clearly, self-betrayal is typically accompanied by profound regret, by the agonized retrospective question, "What have I done?" The problem with making regret a suffi-

9 "Ambivalence as such entails a mode of self-betrayal. It consists in a vacillation or opposition within the self which guarantees that one volitional element will be opposed by another, so that the person cannot avoid acting against himself. Thus, ambivalence is an enemy of self-respect" (Frankfurt 1999e, p. 139; see also Frankfurt 1999, ch. 8).

cient or necessary condition for a departure from past commitments to count as self-betrayal is that there is no reason to think that feelings of regret or shame should deserve automatic authority. Just as second-order desires can be corrected by third-order desires, one can regret feeling regret about something. Put slightly differently, this approach leaves unanswered the question of why wholehearted regret should necessarily trump the presumed wholeheartedness of one's departure from past loves.

These difficulties are a quite general feature of Frankfurt's approach, as long as he remains committed to his volitionalist conception of the authority of desires. His view is that the wholeheartedness of one's caring about something made the associated actions authoritative, and thus that one can trust that when one feels no pangs of conscience or feelings of internal conflict, then there was no reason to doubt the authority of one's commitments. But, as we have just seen, this rules out providing an acceptable account of self-betrayal. We need a more robust account of how one's caring about something can get a normative grip on the deliberating agent.

3. An inferentialist understanding of wholeheartedness

One suggestion is to construe the sort of internal connectedness characteristic of wholehearted agents as including not simply a strong positive feeling toward one's involvement in the project (combined with no misgivings), but also a package of activities and emotional responses that are constitutive of meeting the standard of "genuinely caring about something." Bennett Helm has recently suggested, for example, that just as one's emotional responses must connect together to form a coherent pattern if they are to count as an emotion at all, caring wholeheartedly has similar implications:

[T]o care about some goal is not only to want it but also (other things being equal) to be afraid when its accomplishment is threatened, to be hopeful when it might well be achieved, to be angry at those who impede one's progress, to be frustrated at repeated failures, etc. These various emotions thus converge on a common object (and would consistently continue to do so in the relevant counterfactual situations), and it is in this way that they form a projectible pattern (Helm 1996, p. 76-7).

Although presented as a departure from and corrective to Frankfurt's approach, at least this aspect of Helm's proposal is a natural extension of Frankfurt's discussion. It helps to make clear the requisite intrapersonal connections among one's actions, one's responses to situations, and one's desires, that must be in place if one is to be viewed as acting from desires that are one's own and that have authority. Because it engages a whole array of affective, cognitive, and behavioral dispositions, it generates an account of authoritative desires that is more deeply and broadly anchored in the agent. By focussing on a wide range of elements that must cohere internally, this approach offers a broader base for making claims both about what the holistic pattern is with which one's choices cohere and about how much has to change for the departure from one's prior love for a person, project, or ideal to count as an

authentic change of heart.¹⁰

The attempt to develop a plausible intrasubjective account might draw on the “inferentialist” idea that a desire gets its authority by fitting into a network of connections. Elijah Millgram, for example, has developed an inferentialist approach that distinguishes a desire (as something that has authority or can appropriately play a role in practical reasoning) from other states, such as urges, wishes, and whims (Millgram 1997, ch. 2). What makes a state a genuine desire – in the language I am using, what gives it authority – is the degree to which the desiring agent lives up to *inferential commitments* entailed by the having of such a desire. Helm’s talk of emotional patterns is a case of this: my caring about my family commits me to being distressed if they are harmed. But it also may commit me to buying them gifts on their birthdays and to not speaking ill of them.

In passing, it is important to note that desires and their inferential commitments are coeval, on this model. It is not that I incur various obligations to help a friend when I become that person’s friend, the way becoming a father legally commits me to provide for my son. For such a model presupposes that we can first identify the desire and then see what follows from it. Rather, the inferentialist point is that my having an authoritative desire is a function of being in a network of affective, behavioral, and cognitive commitments. On this structural point, Frankfurt is in agreement with inferentialists.¹¹

As Millgram points out, these inferential commitments require not only synchronic integration but *diachronic* integration as well (Millgram 1997, p. 64). His own focus is on the way in which the commitments to future action that inferentially underwrite one’s having a genuine desire must also fit with one’s inferential commitment regarding how this desire arose in the first place. He argues that some degree of fit between the “forward-directed inferential commitments” and the “backward-directed inferential commitments” is itself necessary for a mental state to qualify as a desire (again, as distinct from a wish, urge, whim, etc.) and thus as reason-giving.¹² But for purposes of understanding self-betrayal, the key point is the point already introduced by Frankfurt, namely, that acting in certain ways in the future is constitutive for genuinely having the desire in question. For example, if someone claims that he wants to help more with the laundry but never does so,

10 There are several aspects of Helm’s approach that depart rather markedly from Frankfurt, especially in his emphasis on the “rationality” of patterns of emotional response (see especially Helm 2001). In his discussion of what makes a pattern of emotions and desires “rational,” Helm moves toward a position that presupposes an agent-independent understanding of what makes a pattern appropriate and away from an account that makes it a function of whatever pattern one could actually will. The further he goes in that direction, the greater his distance from Frankfurt. In this regard, Richard Double’s talk of “Individual Management Style” (1992) is a pattern-based approach that, although lacking Frankfurt’s insight into reflexivity, is much closer to Frankfurt’s approach in terms of the attempt to be value-neutral.

11 Millgram’s own approach, I should emphasize, is not volitionalist.

12 There are interesting parallels between Millgram’s discussion of the unsustainability of desires that one discovers are caused – e.g., simply by having taken a pill (1997, ch. 2) – and the discussion of the “historical” constraints on a desire being autonomous, found in Christman (1991 and 1993); Mele (1993 and 1995); and Fischer and Ravizza (1998), ch. 7. There are also parallels with Ekstrom’s coherentist approach (1993).

there are good reasons to say – in the absence of any further explanation – that he is mistaken about his claim about what his desires are.

4. Picking the patterns that matter

Although the shift to a broader, inferentialist network of connections does make more plausible the intrapersonal approach to explaining how a current desire or commitment can get the sort of grip on future behavior that supports accusations of self-betrayal, it still does not solve the problem we faced earlier regarding distinguishing self-betrayal from a change of heart or a previously neglected pattern. For the case just mentioned – and many more like it – leave open the question of whether the behavior speaks louder than words and thus establishes the authority. One's actions may *believe* one's words, but unless we know that one's avowed intentions articulate what really matters to one, there is no basis for charging one with self-betrayal.

The problem again is with keeping open the possibility for change without rendering the concept of self-betrayal empty (which would happen if one saw every failure to live up to forward-directed inferential commitments as showing that one never had the desire to begin with). As long as we are not talking about reasons but only patterns or regularities that are internally consistent, the question of *which* patterns are appropriate remains unanswered, and thus it is unclear what the basis might be for charges of self-betrayal. Robert Brandom makes this point by drawing on Wittgenstein's gerrymandering argument:

There simply is no such thing as *the* pattern or regularity exhibited by a stretch of past behavior, which can be appealed to in judging some candidate bit of future behavior as regular or irregular, and hence, on this line, as correct or incorrect. For the simple regularist's identification of impropriety with irregularity to get a grip, it must be supplemented with some way of picking out, as somehow *privileged*, some out of all the regularities exhibited (Brandom 1994, p. 28).

Thus, in almost any of the complex real-world cases in which one's actions could be seen as a departure from a pattern, there are usually *several* possible patterns into which this behavior might fit.

In the modified case of Lord Fawn, for example, his newfound willingness to speak to the steward might be continuous with broader commitments, for example, to not being squeamish in the pursuit of the truth, especially in matters of honor. Some of these patterns will hang together better than others, of course, and this does reduce the indeterminacy. But most of the cases will involve *self-conceptions* playing a decisive role, as the sorts of organizing principles that serve to privilege some patterns rather than others (see, e.g., Korsgaard 1996, pp. 100-7). But note that "*how one thinks of oneself*" is not the sort of thing that can count as part of the pattern in any straightforward sense. For it is already a reflection on the patterns, a principle for giving order to the pattern of one's behavior, emotions, and so on. I shall return to this point below.

It is important to note that Brandom's gerrymandering worry cannot be avoided by extending the temporal scope of our consideration of the pattern. It might be argued that "time will tell" whether the departure from past behavior turns out to be a self-betraying deviation or a part of a larger pattern, which itself has coherence. Thus, it might be argued, we don't need to give up the idea that the standard of self-betrayal is intrapersonal, as long as we understand the scope as appropriately extended. But this does not avoid the problem, for the issue of indeterminacy resurfaces with regard to the question of what counts as the appropriate scope. Seen from close-up, it may be reasonable to consider an action to be a case of self-betrayal. But the further one steps back – the more, that is, one situates the behavior in a new pattern that continues – the harder it becomes to see the departure from prior commitments as anything other than an authentic change of heart.

But even if we *could* say clearly that there was a departure from the relevant pattern, we still face the second issue of how we are to know whether the departure is problematic. Even if, for example, Lord Fawn's willingness to discuss intimate matters with the hired hands constituted a clear departure from prior commitments, it might be a change of heart rather than a case of self-betrayal. To answer this question, we need a way of determining whether the original pattern has current *authority*. We need an account of the sense in which an underlying regularity or pattern in the agent's behavior and dispositional responses can get any normative *grip* on – and thus authority over – what he simply happens to feel and do. The problem is that wholeheartedly caring about a new project, person, or ideal does not explain why it should be privileged, if the previous project, person, or ideal is *itself* something about which one cared wholeheartedly. Why, in the absence of any further explanation, should the *present* wholeheartedness of the caring necessarily trump *past* wholeheartedness, or vice versa? Of course, there usually will be an explanation, but – and this is my point – that will be an explanation in terms of reasons rather than a fit with a pattern, and this takes us beyond the idea that internal coherence can do the job alone.

What is needed, then, is a better understanding of normativity, and of the way in which normativity gets its grip by transcending the subject. The *intrapersonal* approach cannot appeal to the sort of considerations that will explain how commitments get a grip on individuals, considerations that necessarily go beyond what the agent happens to feel is unproblematic.

What is needed, I would like to suggest, is the requirement that one's desires, actions, emotional responses, and so on fit together *in a meaningful way*, that is, in a way that meets intersubjective standards of intelligibility. This further constraint, I am suggesting, provides the most promising way to locate the appropriate *grip* that authoritative desires are supposed to have on one's deliberations.

5. Non-volitional constraints on intelligibility

In a certain sense, of course, any coherent volitional pattern "makes sense" and any wholehearted love for something is "meaningful." But those who follow Frankfurt's volitionalist line tend not to appreciate adequately the extent to which

these impressions of intelligibility are generated at least in part by non-volitional considerations.¹³ In particular, what makes sense is in large part a matter of what the evaluative and conceptual framework of one's community allows. And this is decidedly not a matter of the individual will or even of individual values.

One of the most compelling arguments for this point is found in the work of Charles Taylor. As he puts it, it would be "crazy" to claim that "people could determine what is significant, either by decision or perhaps unwittingly and unwillingly by just feeling that way....I couldn't just *decide* that the most significant action is wiggling my toes in warm mud" (1991, p. 36). The reason for this is that the constraints of intelligibility are part and parcel of the "horizon of significance" within which we always already find ourselves, and which we cannot simply construct ourselves.

When we come to understand what it is to define ourselves..., we see that we have to take as background some sense of what is significant. Defining myself means finding what is significant in my difference from others. I may be the only person with exactly 3,732 hairs on my head, or be exactly the same height as some tree on the Siberian plain, but so what? If I begin to say that I define myself by my ability to articulate important truths, or play the Hammerklavier like no one else, or revive the tradition of my ancestors, then we are in the domain of recognizable self-definitions (*ibid*).

In light of this, we can understand why an attempt to see toe-wiggling as an authoritative commitment is open to criticism on the grounds of intelligibility:

Without a special explanation, this is not an intelligible claim (like the 3,732 hairs above).... What could someone *mean* who said this? But if it makes sense only with an explanation (perhaps mud is the element of the world spirit which you contact with your toes), it is open to criticism....Your feeling a certain way can never be sufficient grounds for respecting your position, because your feeling can't *determine* what is significant (*ibid*).

The explanation for this, Taylor argues, is that we must have a vocabulary of values in which to articulate what is meaningful about a given choice. As Taylor puts this, "Things take on importance against a background of intelligibility" (1991, p. 37). Since, given the interpretive character of practical identity, the very possibility of making sense of one's individual responses to the world presupposes the availability of the appropriate semantic resources, including terms ineluctably evaluative in character, the availability of this language is not some optional extra but is rather an essential component of having an identity at all.

13 Breakdowns of intelligibility often seem like volitional breakdowns, as in disoriented exclamations of "What am I doing?!" (Velleman 1989, esp. ch. 1). But the disruption such lapses of intelligibility bring with them is rather different from the volitional breakdown generated by being confronted with an "alien" desire that one must resist or embrace. Instead, the grounds of one's action are alien to one in a rather different sense of not being recognizable as familiar. They are, as it were, *strangers*. I have argued at greater length for this position in Anderson (1996b), ch. 3.

What I am as a self, my identity, is essentially defined by the way things have significance for me, and the issue of my identity is worked out, only through a language of interpretation which I have come to accept as a valid articulation of these issues. To ask what a person is, in abstraction from his or her self-interpretations, is to ask a fundamentally misguided question... (Taylor 1989, p. 34).

Thus, if personal commitments have authority in virtue of the importance that they take on, there seems to be no getting around the need for a socially shared background or horizon within which this enterprise of guiding oneself can occur.

Taylor's talk of working out one's identity brings us back to the issue of one's self-conception. Earlier, I suggested that one of the things that might help to distinguish an authentic volitional pattern from self-betrayal is how one thinks of oneself. The more one thinks of oneself as "Anton's father" or "a Christian Democrat" or "Susan's dissertation director," the more deeply anchored the corresponding patterns become. Of course, we cannot simply give ourselves these identities by *deciding* to be that. In that sense, Frankfurt is right to say that developing a wholehearted practical identity "is not a matter of telling stories about our lives" (1999e, p. 101). But the self-conception does serve to constrain further what can count as an authentic choice, and in two ways. First, it becomes another element in the inferential network in virtue of which we can be said to have a practical identity in the first place, thereby raising the stakes regarding potential betrayal, for one would no longer be able to appeal sincerely to that self-conception in saying who one is. And second, because self-conceptions are especially dependent on the semantic resources discussed by Taylor, the issue of what can count as an authentic self-conception – particularly in contexts of distinguishing self-transformation from self-betrayal – is particularly subject to constraints of intelligibility.

6. The social normativity of authoritative commitments

Considerations of intelligibility add an additional component to the normative grip that personal commitments can have by further restricting the range of what changes in one's personal commitments can count as authentic and authoritative. It thereby provides further resources, beyond what is available to Frankfurt's exclusively volitional considerations, for saying when a change counts as self-betrayal. But in talking about standards of intelligibility, it is important to avoid arguing for these standards in a way that loses sight of Frankfurt's key insight: that the authority that personal commitments have on a person cannot be reduced to claims about the *worth* of what one cares about. And there is some danger of this in Taylor's approach talk of transcendent sources of value. But since a review of these weaknesses would lead us even farther away from my central concerns here, I shall instead sketch an alternative way of situating these constraints on intelligibility, one that ties them primarily to how norms of intelligibility are built into social practices and to the

pragmatics of how those norms are contested in an ongoing fashion.¹⁴

On the pragmatic, Kantian-constructivist approach I have in mind, standards of intelligibility are best understood as built into social practices that we cannot help caring about, to borrow a phrase from Frankfurt. Norms of intelligibility get their grip on us indirectly, by being embedded in and presupposed by social practices to which we are almost unavoidably committed. Indeed, if we are to have any understanding of what patterns count as appropriate – or how past personal commitments can have authority over us – we *must* share to some extent the attitudes that construct the social world within which ascriptions can occur. This then provides the basis for saying how the intrapersonal approach's requirement that desires fit into a wholehearted pattern to be authoritative gets expanded to include an *intersubjective* component. The proposal then is to take the inferentialist point – about desires having authority (and thus being able to function as reasons) in virtue of their connections to other commitments, desires, dispositions, etc. – and combine this inferentialism with a recognition of the way in which what counts as an *intelligible* connection is constrained by the normative standards built into our shared social practices. The reason-giving capacity of (some but not all) desires is a function of those desires fitting into a network of inferential commitments, a network that itself must be situated in the public "space of reasons." To return to Taylor's example of wiggling one's toes in the mud, this is just what is missing: no matter how much everything coheres internally and how *invested* one may be in the activity, it doesn't make sense to give it the status of a personal commitment that one could *betray*.

One good example of this line of thinking – that is, the idea that we are being subject to normative constraints that are built into social practices that are of our own making – can be found in Peter Strawson's discussion of responsibility (Strawson 1974; cf. Wallace 1994). On his view, we have various deeply sedimented (though sometimes conflicted) practices of taking "personal attitudes" such as resentment toward one another, and these attitudes, taken together, construct a social world in which some individuals are to be held accountable for their actions and others not. The fact that these attitudes could have been different – and that, if they were, responsibility and blame would be fundamentally different – does not make it "up to us" to decide who is responsible and who is not. It is the social world constructed by these attitudes that has a normative grip on us. That grip can be challenged, but it cannot be wished away by individuals (see Brandom 1994).

The implications of this approach for my modified version of Frankfurt's case of Lord Fawn are as follows. Imagine Lord Fawn insisting that his willingness to discuss with the estate-steward his fiancée's philanderings was not a matter of

14 With regard to the "objectivist" difficulties in Taylor's account, see Anderson (1994 and 1996a). Similar issues could be raised in connection with Susan Wolf's discussion of constraints on autonomy having to do with an agent's "sanity" or "ability to act in accordance with the True and the Good" (Wolf 1990, ch. 4 and 6), with Paul Benson's early work (e.g., 1987; for a different view, see Benson [forthcoming]), or with Sigurdur Kristinsson's substantivist approach (2000). Thus, although I agree with Taylor, Wolf, Benson, Kristinsson, and others in their emphasis on there being non-volitional constraints on what one can autonomously will, I have doubts about their metaethical presuppositions.

self-betrayal at all but rather fit perfectly well with the pattern of what he really cared about, if one just understood his commitments broadly enough. After all, not only did he actually go through with it – evidence that it was not “unthinkable” for him – he has had to do things like this in other contexts as well, quasi-legal proceeding for example. What my focus on norms of intelligibility highlights, however, is that there will be constraints on what sorts of evidence will count and what won’t. Arguably, one consideration that *would not* count is, “After all, I already discuss with him the mating activities of the livestock on the estate.”

This wouldn’t count, because no one to whom he would be trying to explain his constancy would see this as salient to the question of what his principles are in discussing intimate human relations. Indeed, if he is *that* confused about the standards of apt conduct, it might appear that there is no genuine pattern there at all. In this way, constraints of intelligibility would serve to narrow the scope of the indeterminacy about whether Lord Fawn was betraying himself. In general, the question of whether neglecting a particular project, relationship, or ideal that someone once cared about is to count as self-betrayal or just a change of heart has to be answered in part on the basis of whether the account he could give (for why it isn’t self-betrayal) meets publicly shared standards of an acceptable answer. Lord Fawn’s sense of what the relevant pattern is, for example, is appropriate only if there are good reasons for picking it out as relevant. No matter how wholehearted his feelings in the matter, the feelings do not decide what will count as a reason. That’s not the way reasons work.

Two points need to be noted. First, I said that this would “arguably” not count as a reason, and I meant that quite literally. I am not making any claim about what *must* count as a good reason or not. One of the advantages of my approach is that it makes the question of what can count as a reason a function of what norms have force within the relevant social practice and contexts. Thus, it is no objection to point out that we can imagine contexts in which drawing the parallel with discussions of animal breeding *would* count as a reason. It is one of the advantages of the view I am proposing that it is sensitive to these contextual and contingent differences regarding what counts as a reason and how much of a reason one is expected to be able to provide, even in principle. For example, it may be part of the contingent (yet firmly rooted) social practices of Western liberal societies that little explanation is required to say why dropping a hobby is a genuine change of heart but that one would need to do *much more* convincing to show that switching one’s basic political allegiances is a matter of shifting interests rather than self-betrayal.¹⁵ One of the distinctive features of liberal cultures is the high degree of discontinuity in one’s attachments to persons, projects, and ideals that is viewed as acceptable – or rather, as counting as an authentic change of heart. This may be something we wish to consider a great achievement of liberalism, but it would

15 Similarly, one could appeal to this idea of a burden of proof to account for intuitions about the asymmetry between how much less explaining is expected when the purportedly authentic change of heart is viewed as a *welcome* development, as compared with a change for the worse. Such a constructivist account would differ markedly from Susan Wolf’s analysis of similar asymmetries; see Wolf (1990), pp. 79-81; and Wolf (1980).

be a mistake to think that liberalism is tracking some truth about the amount of continuity that is “really” required.

Second, it is not my claim that the appeal to social normativity will entirely eliminate questions about whether one is “really” betraying oneself, continuing a pattern, or undergoing an authentic change of heart. In many cases, some degree of indeterminacy will remain. But the advantage of the social normativity is that it takes the issue of indeterminacy out of the domain of metaphysics and relocates it in the practical contexts in which indeterminacy does – or does not – matter. In particular, it directs our attention to the norms governing where the burden of proof lies.

To illustrate this last point, imagine that Lord Fawn acknowledges that although there is no reason *now* to see his willingness to speak to the steward as anything but self-betrayal, *in thirty years* it will seem to fit into a pattern. Whether this is an adequate response is not, it seems to me, a matter of metaphysics. Rather, it is a matter of the norms that happen to be in place within the community in which one is trying to make the case. For some issues and in some contexts, this sort of promissory note will be enough to permit the claim to stand. For other purposes, the relevant norms may dictate that self-betrayal is to be determined on the basis of a more restricted temporal scope.

Making this appeal to norms governing the attribution of the intelligibility and authenticity of volitional patterns raises concerns from two directions: that this approach gives too much weight to conventionalism and that it ultimately leaves everything up for grabs.

In response to the charge that this social approach makes an individual’s autonomy, authenticity, or integrity hostage to social acceptability, I would emphasize three points. First, because the relevant norms have their grip only in virtue of our buying into the practices in which they are embedded, they can always be contested by those who wish to alter the practices. On this approach, the individual is viewed as being a participant in the social world, and thus as continuing to reproduce and modify the practices that support the normative constraints to which she is subject. What has to be acknowledged, then, is the *interplay* between sedimented norms and the individual’s prerogative to challenge them. It cannot be determined in advance how this will work out. It has to *be* worked out, in response to each challenge, in the actual contexts.

Second, the approach is not conventionalist in the sense of excluding iconoclasts, eccentrics, and rebels about specific norms. For it can always be part of the normative structure of our social practices that eccentricities enjoy a degree of acceptability, when they are understood, specifically, as eccentricities rather than serious and incoherent lapses of competence. To illustrate the general point here, take a mundane example of how deviation from norms built into greeting practices can be understood.¹⁶ I know a man who consistently greets people with an Iroquois (American Indian) handshake, which involves grasping the other’s forearm rather than his or her hand. For those unfamiliar with the practice the experience of shaking his hand (so to speak) is decidedly odd, and his deviation from

¹⁶ For a fascinating discussion of the normativity of greeting practices, see Heritage (1984), pp. 106-20.

the standard practice of handshaking generates a great deal of confusion. One initial response is to think that he is not actually greeting you but grabbing your arm, and thus that he is either “up to something”. But then it turns out both that there is a subculture in the United States (a loose-knit community known as the “Rainbow People”) in which this is the appropriate form of greeting, and that there are good reasons for adopting this practice, in that it dramatically decreases the possibility of transmitting disease, while also demonstrating symbolic solidarity with Native American victims of genocide. As a result, this greeting behavior comes to count as eccentric, but acceptably so. Again, the boundaries of the acceptable have to be worked out, but this highlights the way in which a constructivist, contextualist approach to social normativity can accommodate the sorts of deviations that must be accommodated by a commitment to reasonable pluralism.

Third, and relatedly, it is important to stress that my view is *not* that the authenticity of a person’s choices is determined by what the majority of people think. The objection frequently raised against any social conception of normativity is that it would allow people to be labeled crazy by popular consensus even if they really weren’t – just as perfectly sane women were labeled “witches” in the past. But such objections face two serious problems. First, most such objections get off the ground only by assuming that the individuals deemed crazy in fact have good reasons that are not understood by others. Such thought-experiments are precooked in a way that presupposes a standpoint that is not available to situated agents. Consider, instead, our landing among a community of vastly more intelligent, empathic, self-critical beings and having constant trouble providing satisfactory accounts of our choices. What could we then appeal to in defense of our claim that we are actually quite competent in defending our choices? Indeed, the very construction of the thought-experiment relies on the mistaken assumption that there is a perspective from which one can decide in advance whose standards for autonomy are to be preferred. Second, it is important to keep in mind that neither convincing others nor failing to do so can serve as a *criterion* for justification (Brandom, 1994, ch. 9; Habermas, 1999). Since justification is not limited to the conventional, there is always the fallibilistic hope that one’s own understanding will win out. Individuals may come to be seen as more autonomous than they were thought to be earlier. This will be true even in cases where withholding the ascription of autonomy was perfectly warranted, though most such cases – such as European men treating women or Africans as incompetent – can be handled by showing that even by the hegemonic standards of that context, there was little warrant for this view.

This leaves us, finally, with the worry that by moving toward a contextualist and pragmatist view that is accommodating of eccentricity, I have given up the advertised advantage of the whole approach, namely, that it would be able to provide additional normative traction regarding the authenticity of our personal commitments, as compared with Frankfurt’s volitionalist approach. For it now starts to seem as if the force of these norms is really up to us. But this would be to overlook two crucial points. The first is an insight that Brandom traces to Kant:

Although the status of being obliged to follow a particular rule is instituted by our attitudes, what is correct according to that rule is not simply determined by what we *take* to be correct according to it. The status of correctness of a performance according to a rule does not collapse into the attitude of assessing that performance as correct. Endorsing a rule gives it a grip on us. Part of that grip is that the rule does not mean just whatever we later might take it to mean (Brandom 1994, p. 52).

The second point takes us back to the Strawsonian idea that some of our practices, attitudes, and categories are so closely woven into the fabric of our social world that we would be profoundly disoriented without them. It is, of course, conceivable that we could dispense with the practices built around the attribution of authenticity, integrity, and self-betrayal. But it would be a massive change. These concepts are at the heart of our attempts to determine who we are and wish to continue to be (or not) and whom to trust and love (or not). And furthermore, there is an indirect argument of a very Frankfurtian sort for the importance of these practices. They offer us more of the limits that we need for our practical identities to have definition, depth, and therefore authority. If we were able to give up these practices, we would have shaken off the hold of the norms governing when desires have authority, leaving us more rudderless than we already are.¹⁷

7. Conclusion

From the outset it has been clear that we can explain the authority of personal commitments – and thus the phenomenon of self-betrayal – only if one is subject to standards *of one's own* that nonetheless have a normative grip on one's future choices. As Frankfurt emphasized, this further requires that there be a gap between what one is actually inclined to do and what one can want to do without betraying oneself. On the intrapersonal view, coherent intrapersonal patterns of volitions, desires, and emotions were thought to do the job. What was supposed to generate a gap between the given and the authoritative was to be volitional or dispositional. This intrapersonal model aimed to show that the grip could be offered by something deeper or at least something that is internal and entirely structural, something that is sedimented in the will, as it were.

The difficulties I highlighted there suggested that instead what is needed to explain how our personal commitments can get a normative grip on our actual desires is at least indirectly social. The questions about whether the individual's inferential commitments line up in the right way are questions that require the use of judgment and thus justification, which is only to be understood in terms of open procedures of the giving and taking of reasons. Whether a given inferential pattern constitutes a reason is not for me to decide – only for me to argue for. This is why the question of whether a desire is authoritative is ultimately a question of the sort of account that can be given in defense of it fitting in, in response

17 As we have seen, the need for limits is very much a concern of Frankfurt's; see esp. (1999), ch. 9. Where he goes wrong, in my view, is in thinking of the constraints in terms of the volitional essence of a person rather than in terms of reasons.

to challenges raised and so on. This is social in the sense that, although only I can live my life, others can show me to be wrong about my belief that I am experiencing a change of heart rather than self-betrayal.

Specifically, my proposal is that we think about what allows personal values and commitments to work as normative standards is that they could be defended with good reasons as being appropriate for that individual. And here the standards that transcend the individuals are to be found sedimented in social practices rather than in the depths of the will.

The surprising result of this discussion is that being autonomous and preserving one's integrity should require so much attention to what others say. But this is just another way of formulating the not so surprising result, namely, that if the standards we endorse are to have a grip on us that goes beyond the fact that we happen to want currently to act in accordance with them, then we must hold ourselves accountable to something that we endorse and yet do not have at our disposal. And the best candidate for that are public processes of the giving and taking of reasons.¹⁸

CV

Joel Anderson is Senior Lecturer in Philosophy and Associate Director of Social Thought & Analysis at Washington University in St. Louis. He works in the areas of ethical theory (especially on autonomy and discourse ethics), applied ethics (especially on issues in 'neuro'-ethics), and social theory (especially on the role of relations of mutual recognition in securing conditions for acquiring autonomy and other fundamental human capabilities). He has edited, with John Christman, *Autonomy and the Challenges to Liberalism: New Essays* (forthcoming).

References

- Anderson, J. (1994) "Wünsche zweiter Ordnung, starke Wertungen und intersubjektive Kritik: Zum Begriff ethische Autonomie." *Deutsche Zeitschrift für Philosophie* 42: 97-119.
- Anderson, J. (1996a), "The Personal Lives of Strong Evaluators: Identity, Pluralism, and Ontology in Charles Taylor's Value Theory", *Constellations: An International Journal of Critical and Democratic Theory* 3: 17-38.
- Anderson, J. (1996b), "A Social Conception of Personal Autonomy: Volitional Identity, Strong Evaluation, and Intersubjective Accountability", Ph.D. dissertation, Northwestern University.
- Benson, P. (1987), "Freedom and Value" *Journal of Philosophy* 84: 465-86.
- Benson, P. (forthcoming), "Taking Ownership: Authority and Voice in Autonomous Agency", in *Autonomy and the Challenges to Liberalism: New Essays*, J. Christman and J. Anderson (eds.), ms. under review.
- Bittner, R. (1992), "Is it Reasonable to Regret The Things One Did?" *Journal of Philosophy* 89: 262-73.

18 This essay was initially written for a conference on "Handlungspielräume" at the Institute for the Cultural Sciences in Essen, Germany, organized by Martin Löw-Beer and Gertrud Koch. I am grateful for comments there from John Christman, Ulrike Heuer, Matthias Kettner, Susan Nieman, Beate Rössler, Martin Saar, Holmer Steinfath, Gary Watson, and Rolf Wiggershaus. For further comments on and discussion of earlier drafts, I would also like to thank Angela Bolte, Eric Brown, Larry Davis, Marilyn Friedman, Pauline Kleingeld, Sigurdur Kristinsson, Larry May, Thaddeus Metz, Brooks Siegel, Eleonore Stump, and Carl Wellman. Two anonymous referees for *Philosophical Explorations*, also made extremely helpful suggestions. This large number of interlocutors does nothing, however, to reduce my responsibility for my assertions here.

- Bittner, R. (2002): "Autonomy, and Then", *Philosophical Explorations* 5: 217-28.
- Brandon, R. (1994), *Making It Explicit: Reasoning, Representing, and Discursive Commitment*, Cambridge, Mass.: Harvard University Press.
- Brandt, R. (1979), *A Theory of the Right and the Good*, Oxford: Oxford University Press.
- Burge, T. (1979), "Individualism and the Mental," *Midwest Studies in Philosophy*, 4: 73-121.
- Buss, S. (1994), "Autonomy Reconsidered," *Midwest Studies in Philosophy* 19: 95-121.
- Christman, J. (1991), "Autonomy and Personal History", *Canadian Journal of Philosophy* 21: 1-24.
- Christman, J. (1993), "Defending Historical Autonomy: A Reply to Professor Mele", *Canadian Journal of Philosophy* 23: 281-90.
- Christman, J. (2001), "Liberalism, Autonomy, and Self-Transformation", *Social Theory and Practice* 27.
- Dennett, D. (1984), *Elbow Room: The Varieties of Free Will Worth Wanting*, Cambridge, Mass.: MIT Press.
- Double, R. (1992), "Two Types of Autonomy Accounts" *Canadian Journal of Philosophy* 22: 65-80.
- Ekstrom, L. (1993), "A Coherence Theory of Autonomy", *Philosophy and Phenomenological Research* 53: 599-616.
- Fischer, J. M. and M. Ravizza (1998), *Responsibility and Control: A Theory of Moral Responsibility*, New York: Cambridge University Press.
- Frankfurt, H. (1987), *The Importance of What We Care About*, New York: Cambridge University Press.
- Frankfurt, H. (1999), *Necessity, Volition, and Love*, New York: Cambridge University Press.
- Frankfurt, H. (2000a), "The Dear Self", *The Philosopher's Imprint* 1: 1-14
www.philosophersimprint.org/001001>
- Frankfurt, H. (2000b), *Some Mysteries of Love*, The Lindley Lecture. Lawrence: University of Kansas Press.
- Friedman, M. (1986), "Autonomy and the Split-Level Self", *Southern Journal of Philosophy*, 24: 19-35.
- Friedman, M. (2000), "Autonomy, Social Disruption, and Women" in *Relational Autonomy: Feminist Perspectives on Autonomy, Agency, and the Social Self*, MacKenzie and Stoljar (eds.) New York: Oxford University Press, pp. 35-51.
- Habermas, J. (1999), "Richtigkeit versus Wahrheit. Zum Sinn der Sollgeltung moralischer Urteile und Normen", in *Wahrheit und Rechtfertigung*, Frankfurt: Suhrkamp, pp. 271-318.
- Helm, B. (1996), "Freedom of the Heart," *Pacific Philosophical Quarterly* 77: 71-87.
- Helm, B. (2001), *Emotional Reason: Deliberation, Motivation, and the Nature of Value*. Cambridge: Cambridge University Press.
- Heritage, J. (1984), *Garfinkel and Ethnomethodology*, Cambridge: Polity Press.
- Korsgaard, C. (1996), *Sources of Normativity* (New York: Cambridge University Press, 1996).
- Kristinsson, S. (2000), "The Limits of Neutrality: Toward a Weakly Substantive Account of Autonomy", *Canadian Journal of Philosophy* 30: 257-86.
- Mackenzie, C. (2002), "Critical Reflection, Self-Knowledge, and the Emotions", *Philosophical Explorations* 5: 217-28.
- Mele, A. (1991), "History and Personal Autonomy" *Canadian Journal of Philosophy* 23: 271-80.
- Mele, A. (1995), *Autonomous Agents: From Self-Control to Autonomy*. New York: Oxford Univ. Press.
- Millgram, E. (1997), *Practical Induction*, Cambridge, Mass.: Harvard University Press.
- Noggle, R. (1997), "The Public Conception of Autonomy and Critical Reflection", *The Southern Journal of Philosophy* 35: 495-515.
- Strawson, P. (1974), "Freedom and Responsibility," in *Freedom and Resentment and Other Essays*, London: Methuen, 1-25.
- Taylor, Ch. (1989), *Sources of the Self: The Making of Modern Identity*. Cambridge, MA: Harvard University Press.
- Taylor, Ch. (1991), *The Ethics of Authenticity* Cambridge, MA: Harvard University Press.
- Thalberg, I. (1978), "Hierarchical Analyses of Unfree Action", *Canadian Journal of Philosophy* 8: 211-26.
- Velleman, J. D. (1989), *Practical Reflection*, Princeton: Princeton University Press.
- Wallace, R.J. (1994), *Responsibility and the Moral Sentiments*, Cambridge, Mass.: Harvard Univ. Press.
- Wolf, S. (1980), "Asymmetrical Freedom", *Journal of Philosophy* 77: 151-166.
- Wolf, S. (1990) *Freedom within Reason*. New York: Oxford University Press.