

KNOWING YOUR OWN
STRENGTH:
*Accurate Self-
Assessment as a
Requirement for
Personal Autonomy*

JOEL ANDERSON AND WARREN LUX



ABSTRACT: Autonomy is one of the most contested concepts in philosophy and psychology. Much of the disagreement centers on the form of reflexivity that must have to count as genuinely self-governing. In this essay, we argue that an adequate account of autonomy must include a distinct requirement of *accurate self-assessment*, which has been largely ignored in the philosophical focus on agents' ability to evaluate the desirability of acting on certain impulses or values. In our view, being autonomous (i.e., self-guiding) involves understanding the extent to which one has the capacities required for one's intended actions. On both clinical grounds (drawn from cases of frontal brain injury) and conceptual grounds, we argue that one's autonomy is diminished to the extent to which one's ability to assess one's capacities is impaired.

KEYWORDS: insight into illness, executive function, deficit awareness, frontal head injury, reflexivity, authenticity, compensatory strategies, anosognosia

ACTING AUTONOMOUSLY REQUIRES certain capacities for reflexivity. Debates in the philosophical literature on autonomy often turn on what form of reflexivity is required,

whether it be endorsing the motives on which one acts in "second-order desires" (Dworkin 1988, ch. 1; Frankfurt 1988, ch. 1), "giving the law to oneself" (Kant 1785/1997), or even just caring about a person or project (Frankfurt 1999, ch. 11, 14). For the most part, however, these debates center on issues of practical reasoning and evaluative self-legislation. In the present essay, by contrast, we argue that there is a distinct requirement of *accurate self-assessment*, which is connected more with executing a task than with evaluating its desirability. To act autonomously, we claim, one must understand the extent to which one has the capacities required for one's intended actions. On both conceptual grounds and on clinical grounds (drawn from cases of frontal brain injury), we argue that one's autonomy is diminished to the extent to which one's ability to assess one's capacities is impaired. Although a similar requirement may be implicit in some accounts of autonomy (e.g., as a matter of minimal rationality), we believe that it deserves to be treated as a distinct requirement. Acknowledging the distinct importance of awareness of

deficit is also important for appropriately conceptualizing clinical cases in psychiatry and neurology, especially cases having to do with loss of executive function caused by head injury.

As an illustration of the deficit we have in mind, consider the real clinical case of “John.”¹ A severe frontal injury contused the anterior portions of John’s brain and at the same time shattered both of his orbits, severing his optic nerves and leaving him with no light perception at all. The resulting behavioral syndrome was quite striking in that John not only insisted verbally that he still had vision, but he also initiated behavior as if he did, trying to move about his room in the manner of a person with normal vision. As a result, he walked into walls and furniture, collided with objects in his path rather than avoiding them, and repeatedly placed himself in positions that were extremely precarious for a person who could not see. Despite his ability to initiate action in an apparently self-directed way, John’s persistently mistaken assessment of his visual capacity with respect to his actions made it impossible for him to act as he intended. In this sense, many of his actions could not count as autonomous, not because he could not see—plenty of blind individuals are perfectly autonomous—but rather because his impaired self-assessment left him unable to make sense of what he was doing. At least with respect to those actions, he was deeply alienated from himself as an agent.

The neurologic details of this case are important, however, for although John has a variant of a condition called Anton syndrome, the localization of his lesions differs from that of patients with classical Anton syndrome in ways that are central to our focus here. The essential clinical features of Anton syndrome are blindness and visual anosognosia (lack of awareness of a visual deficit). Individuals with Anton syndrome, like John, are blind but deny that they are blind. In classical Anton syndrome, however, patients have bilateral posterior cerebral lesions that involve primary visual cortex and also extend into nearby visual association areas. In John’s case, on the other hand, his entire posterior brain was unaffected. His blindness was not cortical in origin

but rather resulted from direct trauma to his optic nerves. His unawareness, moreover, was related to extensive damage to both of his frontal lobes that resulted in profound impairment of his executive functions, functions that we discuss in greater detail later in this paper. For John, as for many other patients with frontal lobe injuries, his executive function deficits rendered him unable to integrate a knowledge of any of his impairments, including his blindness, into his behavioral output, verbal or otherwise, at any level. Although we agree that denial of visual loss alone, as seen in the visual anosognosia of classical Anton syndrome, would be sufficient to compromise autonomy, the distinct and more significant threat to John’s autonomy, in our view, relates to how his executive function disorder compromised his overall behavioral output. For even if he had been aware of his visual deficit at some level (comparable, for example, to the limited level of deficit awareness shown by Henry; see *The Case of the Barge Worker*), the damage to John’s frontal capacities for integration of that information would have rendered him unable to use his awareness to direct those aspects of his behavior that depended on vision.²

The category *insight into illness*, commonly used in psychiatric practice, also bears some discussion here. Many forms of psychopathology are associated with diminished insight into illness, and clinical observations related to that fact are not uncommon in psychiatric patients. Whether the responsible mechanisms in many of these instances are psychological or neurologic (or both) remains a subject of controversy. What is clear, however, is that “insight into illness” is a broad category that includes, but is not limited to, the rather specific issues with which we are concerned in this paper. In each of the cases that we cite, the neurologically based deficit in awareness is due to identifiable, localizable brain disease and can be analyzed for its neurocognitive consequences on that basis. In each instance, the deficit in awareness diminishes insight into illness, but our concern in this paper is with a careful analysis of the precise, but limited, neurologic mechanisms and their implications for the nature of personal autonomy. We readily

acknowledge that this may also have implications for understanding the relationship of insight to autonomy more generally, but that is a broader issue.

This paper is divided into five sections. We begin by defining *accurate self-assessment* and clarifying several features of that definition. We then articulate in greater detail the conceptual argument for why accurate self-assessment is central for autonomy. Next we provide some brief neurologic background on the underpinnings of autonomous action, with special focus on the “executive function” accomplished primarily in the prefrontal cortex, which is crucial for understanding the details of a second clinical case that we use to further elaborate, in the following section, our claims for the importance for autonomy of accurate capacity self-assessment. We conclude by responding to several potential objections.

DEFINING “ACCURATE SELF-ASSESSMENT”

Before turning to the neurologic background and a more detailed clinical case, it is useful to clarify conceptually the requisite capability. We say that a person “lacks accurate self-assessment of her capacities” with regard to a given task if (1) she intentionally or deliberately undertakes the task, (2) she has the sense that she or he will be able to perform the task as intended, and (3) the self-assessment of capacity underlying this sense is inaccurate.

Several phrases here are worth explaining briefly. To begin with, “having a sense” of being able to do something is meant broadly. There is no requirement here of explicitly formulated beliefs. Rather, the “sense” that one has of one’s abilities is a complex dispositional property, reflected, for example, in one’s surprise at not being able to do something or one’s reluctance to undertake a task. A figure skater going into a difficult part of a routine, for example, may decide to scale back to a double jump on the basis of her “sense” that she may not be able to pull off the triple jump. What is required, however, is that the agent’s sense of her capacity is a component of her relation to herself as the agent of actions she undertakes—precisely what was lacking in John’s case.

Second, because the concept of “capacity self-assessment” (at least as we are using it here) has a task-relative dimension, it must include an understanding of the level of one’s capacity in relation to the difficulty of the task in question. For this understanding to be “accurate,” therefore, the task parameters themselves must be perceived to an adequate degree of precision. For example, to acquire a sense that one can jump a hurdle successfully, one must accurately estimate not only the strength of one’s legs but also the height of the hurdle. For our analytical purposes here, however, we focus on those faulty self-assessments in which the elements of the task itself have been accurately discerned.³

Third, although we sometimes speak here of “having” or “lacking” the capacity for accurate self-assessment, we are well aware that this is not a binary matter but rather a matter of degree. In fact, there are several continua involved. Not only does the accuracy of one’s self-assessments admit of degrees, those self-assessments are themselves best expressed probabilistically. For example, one’s sense that one can run fast enough to catch a bus or that one has studied enough to pass an examination is normally qualified, with the degree of confidence reflecting the judgment of probability. Relatedly, it is important to underscore that we are not arguing for any particular level of inaccuracy entailing nonautonomy. Given that accuracy of self-assessments is a matter of degree, there will always be an issue of how much accuracy is enough. But this will always be an issue of how much is enough for certain practical purposes. To make our conceptual point that autonomy presupposes accurate self-assessment, however, we need not specify any general thresholds or necessary and sufficient conditions for “accuracy.” It should be clear enough that the cases we discuss involve self-assessments that are sufficiently off the mark to undermine the attribution of autonomy. But for most people and most purposes most of the time, a rough-and-ready sense of one’s capabilities suffices.

Fourth, we are construing “accuracy” along the lines of epistemologic externalism. The accuracy of one’s self-assessment is a function of

whether it corresponds to the facts, not a function of what it is subjectively reasonable for one to believe about one's capabilities. Thus, if one had no way of knowing that one has been incapacitated, it would be subjectively reasonable to assume that one had the relevant capacity, but the inaccuracy of that self-assessment would still generate autonomy-impairing problems for action (as we discuss in the next section). It might seem overly strict to view persons as less autonomous because of factors outside their control, but this worry can be handled easily by pointing to the role of excusing conditions in our practices of autonomy attribution. In cases where agents come to recognize and correct a reasonable but mistaken self-assessment, their overall autonomy is not in question, although we (and they) would still view as not fully autonomous the action undertaken prior to this realization. In the cases in which agents' inaccurate self-assessments persist in the face of available evidence, the self-assessment will count not only as inaccurate but unreasonably inaccurate—typically with accordingly more serious practical implications. As an unfortunate empirical matter, most of the head injuries that impair self-assessment also impair the ability to learn from feedback, thus rendering unreasonable many of such patients' inaccurate self-assessments. As a conceptual matter, however, even subjectively reasonable inaccuracy undermines the autonomy of the relevant action.

Fifth, in restricting ourselves to cases in which the task is "intentionally or deliberately undertaken," we wish to set to one side the issue of whether unintentional action can be autonomous. Thus our claim that inaccurate self-assessments of one's capabilities can undermine one's autonomy applies only to cases in which an agent acts with a conception of her task and of herself as the agent of the task. This form of agentic or executive reflexivity is typically thought to be a precondition for autonomous action, but that is not something we argue for here.⁴ In addition, our focus on tasks that are intentionally undertaken serves to exclude unrealizable fantasies, wishes, and other idle thoughts from counting as evidence of inaccurate self-assessments, because it is constitutive of such mental states that we do

not view their realization as feasible (Anscombe 1985, sec. 36; Millgram 1997, 13).

Sixth, although we focus on cases in which individuals overestimate their capacities, the same conceptual point applies to many cases of underestimation as well. This is clearest when successful execution of an intended task requires striking a balance between exerting oneself too much and not enough, as in cases where someone breaks something because he "did not know his own strength." Something similar might be said for instances in which someone is paralyzed by the failure to recognize that he has a certain capacity (see the section on Objections).

Finally, in our discussion here we limit the targets of self-assessment to physical and cognitive capacities of the individual. We hold open the possibility that similar points could be made about the broader set of resources needed for accomplishing a task. For example, it may be that autonomy is undermined by an inaccurate assessment of one's financial resources. And although we focus on the capacities determined by an agent's one body and mind, we readily acknowledge that these are both artificial and fuzzy lines, perhaps even to the extent suggested by the literature on the "extended mind" and on mental prostheses (Clark and Chalmers 1998; Rovane 1998, ch. 4).

With this definition in place, we now turn to explaining why accurate self-assessment is required for autonomy and to the question of what advantages this requirement has over more commonly used requirements, such as rationality, "reasons-responsiveness," or integration of beliefs.

SELF-ASSESSMENT AND AUTONOMY

One might wonder why accurate self-assessment is needed for autonomy. Although the lack of autonomy might seem obvious in the case of John (or that of Henry; see *The Case of the Barge Worker*), his actions seem to be generally quite self-directed. Moreover, to establish nonautonomy, one must do more than point to the fact that individuals are making poor or imprudent decisions, or that they are not well-adapted to their

environment, or that some of their beliefs are unjustified. Autonomy should allow for at least some mistaken beliefs. Indeed, respecting autonomy is usually understood as being partly a matter of letting people make their own mistakes. The deficit must thus be something that gets in the way of the agent's exercise of self-governance, or there must be something special about the beliefs at stake.

In our view, inaccurate self-assessment undermines the autonomy of one's actions by making those actions less fully one's own. It is widely accepted in the philosophical literature that autonomy is a matter of acting on desires, values, plans, and so on that are "one's own," that is, authentic.⁵ For example, if the process by which those action-orienting mental states are acquired involves distortion or manipulation—such as hypnosis or coercion—then the actions that flow from those mental states are not authentically one's own actions, but are rather more like knocking over a glass because one was shoved from behind. In these paradigm cases of nonautonomous actions—such as threats, hypnosis, or intoxication—the sources of one's action have, in some sense, been hijacked. The agent is no longer autonomous because what is setting the agenda for her action is not herself at all. Obviously, things get very complicated very quickly in trying to explain how something manifestly an internal component of one's motivational system can fail to be "one's own."⁶ For our purposes, however, it is enough to show that, at least on two dominant approaches to the authenticity requirement, agents can fail to meet that requirement owing to inaccurate self-assessments of their capacity. We now consider these two dominant philosophical approaches in turn.

One common way of specifying when a person's actions are not "her own" is in terms of the causal link between the agent's acting and the source of the action. As authors such as Alfred Mele, John Martin Fischer, and Mark Ravizza point out, an action can fail to qualify as "one's own" because the agent is not properly hooked up to the action. For example, if the action is caused in a way that bypasses any possibility of evaluating the motives behind one's actions (Mele

1995) or of engaging one's reasons-responsive mechanism (Fischer and Ravizza 1998), then there has been a short circuit of the pathway by which ownership of an action is established.

Something similar is involved in cases of inaccurate self-assessment. Someone who is seriously confused about her capacities is like someone operating a remote control that does not function as she expects it to. In such cases, there is a disconnect between one's intentions and one's actions. What these individuals intend and what they do have come apart to such a degree that it becomes implausible to say that they are genuinely governing their actions. Someone who is significantly mistaken about what the buttons on his behavioral console actually do is not really self-guiding. Of course, it would be too strong to make it a necessary condition for autonomy that one's actions always come out exactly as intended or that we anticipate all the repercussions of our autonomous acts. But we also need to avoid saying that the self-governing agency of someone with "phantom limb syndrome" is not at all diminished by his being out of touch with whether he is moving a nonexistent limb. Without accurate self-assessment of capacity, agents are not properly connected to themselves as the agents of their actions.⁷

Or consider the example from which the title of our essay is derived: cases in which a person's action is not autonomous because she "didn't know her own strength." This expression is clearly meant as an appeal to an excusing condition: because the agent is out of touch with her strength, her action is out of control. Such attempts to avoid responsibility are often viewed with suspicion, and justifiably so, because most people have every reason to know how strong they are. But as we have said earlier, that is an issue of whether the mistakes in one's self-assessment are reasonable—of whether one should be held responsible for what one ought to have known about one's capacities. But if we imagine a case of someone being given unknowingly a performance-enhancing drug or a bionic implant, it is clear that the harm the person might cause would not be her responsibility. This would decidedly not be because drugs or implants automatically

render the action subject to an “alien cause” (for they do not if their acquisition is voluntary) but rather because the agent is no longer properly hooked up to the action.

A second way of analyzing inauthenticity is in terms of problems of integration. Here, the sense in which the agent fails to be hooked up appropriately to his motivations is that he cannot make sense of them or is alienated from them—or, more precisely, cannot make sense of himself as the agent of actions motivated in that way. The mental states underlying his actions are not his own because they are *outliers*, states that do not fit intelligibly into his plan of action.

One might object that this approach seems to license rationalizations in which we appeal to “unintelligibility” to avoid responsibility for actions that flow from shameful aspects of our character.⁸ This is significantly less of a problem as soon as one realizes that the constraints in question here (namely, the constraints of what “makes sense”) are not under the individual’s control. As has been emphasized in the literature by Elijah Millgram, David Velleman, and Harry Frankfurt, simply deciding to desire something undermines the authority of that desire (Millgram 1997); the rationalization itself is also subject to the demands that it be intelligible (Velleman 1989); and one cannot simply decide what one wholeheartedly cares about (Frankfurt 1988, ch. 12, 1999, ch. 9, 11). One corollary of the principle comprised by these observations is that one needs a certain degree of integration if anything is to count as making sense (or not). And the additional corollary important for our purposes is that one needs a feedback mechanism: one must be constituted in such a way that the unintelligibility surfaces. For to the extent to which one is unable to note the internal tensions, one is without this compass, which is so crucial for guiding one’s actions in the manner we dub “autonomous.”

And this is why rigidly inaccurate self-assessments undermine autonomy. Think, again, of John, who cannot recognize his own blindness. His persistently mistaken self-assessment makes it impossible for him to make sense of why his plans did not work out or why he keeps running

into walls. For there is an element in the generation of his action (the belief that he is sighted) that he cannot possibly square with his experience. Note that this is not just a point about his cognitive rigidity, although rigidity often characterizes such patients. It is more generally a point about the way in which the ability to guide one’s actions in a way one can make sense of requires that the constituents of one’s practical reasoning—one’s desires, beliefs, values, and so on—be sufficiently integrated inferentially for errors to register as errors. This is the point that Bernard Williams makes in his discussion of a man who tries to deliberately bring himself to believe that his dead son is alive:

The man gets rid of this belief about his son, and then there is some belief which strongly implies that his son is dead, and that has to be got rid of. Then there is another belief which could lead his thoughts in the undesired direction, and that has to be got rid of. It might be that a project of this kind tended in the end to involve total destruction of the world of reality, to lead to paranoia. (Williams 1973, 151)

The key point here is that even if the father succeeded in his intentional self-deception, it would be globally self-defeating, for it would wreck havoc on the inferential fabric that makes it possible for him to orient himself in the world. Similarly, one way of diagnosing the autonomy-impairing disorientation suffered by John is to say that his prefrontal injury generates so many “bugs” in the inferential system he relies on in planning and executing even mundane tasks that his plans and intentions are no longer sufficiently integrated into the self to count as being genuinely his.⁹

Our claim, then, is that whichever account of authenticity one favors, inaccurate self-assessment undermines the extent to which an action is one’s own. It might be wondered, however, especially in light of this last line of argument, whether it would not be more straightforward to focus on one of the more familiar requirements found in the autonomy literature, such as a requirement that one be minimally rational (Christman 1991; Dworkin 1988), that one have a “moderately reasons-responsive mechanism” (Fischer and Ravizza 1998), that one be “objective” (Berof-

sky 1995; Wolf 1990), and so on. We do not wish, here, to choose sides in this debate or even to argue against any of these ways of construing the requirements of autonomy. Our point is rather that any defensible set of requirements for autonomy must include (or entail) the capacity for accurate self-assessment and that focusing directly on self-assessment has the threefold advantage of being (1) more neutral vis-à-vis competing theories, (2) more plausibly tied to the active reflexivity constitutive of autonomy, and (3) more directly supported by evidence from clinical neuroscience.

- (1) Because our focus is restricted to self-assessment, it can be adopted without committing oneself to the stronger and more controversial requirements, for example, that one be fully rational, or that one's choices be the right ones, or that one's choices be made in light of full appreciation of all the relevant information about what one is choosing. There are important debates over these issues, and the more demanding requirements may end up being justified. Either way, because being accurate in one's assessment of one's capacity is a more limited requirement, it is easier to take on board.
- (2) Perhaps the most important advantage of our approach is that the component of overall rationality it targets seems particularly central to autonomous action. Calls for a broader, more demanding requirement of rationality can easily get bogged down in trying to defend the full scope of the requirement and often end up having to appeal to tangentially related considerations regarding the instrumental value of autonomy. But the relationship of these justifications to the nature of autonomy is much less direct than in the case of the requirement of accurate self-assessment, which focuses on the very ability to connect to one's intended action with oneself as the agent of the action. What is particularly apt about this requirement is that it highlights the agentic reflexivity that is one of the hallmarks of autonomy.
- (3) Finally, the focus on self-assessment also has the advantage of being able to draw support from clinical neurology. In particular, there seem to be mutually reinforcing intuitions about, on the one hand, a cluster of capacities that constitute an intuitively plausible and conceptually coherent account of autonomy and, on the other hand, a package of neurologic capacities that has been termed "executive function." For, if it

is plausible that the widely used neurologic concept of "executive function" is broadly isomorphic with capacities associated with autonomy, then the conceptual claims made thus far will find corroboration in observations about impaired executive function. And it is to these observations (and the background for them) that we now turn.

EXECUTIVE FUNCTION, SELF-ASSESSMENT, AND TRAUMATIC BRAIN INJURY

The term *executive function* is used by neurologists, neuropsychologists, and other clinical neuroscientists to refer to that set of neurocognitive capacities that are directly engaged when one plans, initiates, and carries out a goal-directed activity over time with appropriate self-monitoring and self-correction as one proceeds. Its critical neural substrate is the anterior forebrain, and it gives cognitive structure and coherence to human action in novel situations. Whereas posterior (perceptual-motor) and basal (emotional) function may be sufficient for behavior in familiar settings, activation of executive function is necessary to guide oneself prospectively in the unstructured real world. Key among the capacities that make up executive function are anticipation, goal selection, planning, initiation, sequencing, monitoring (error detection), and self-correction (initiation of novel responses).¹⁰ Of particular relevance to the thesis of this paper, moreover, is the capacity for self-assessment. The importance of this capacity to executive function derives from the fact that effective task execution is possible only when one has the ability to accurately assess whether one has the physical and mental skills required to carry out an intended act. When the capacity for self-assessment is working well, an agent is able to accurately assess other capacities that she possesses (and is able, in fact, to accurately assess her own capacity for self-assessment as well). When the capacity for self-assessment is not working well, accuracy of some or all of these assessments is impaired. Consider, for example, the case of an individual

who wants to get from point A to point B on the other side of a room but has one paralyzed leg and is unable to walk without assistance. If she cannot accurately assess her ability to bear weight on her leg and as a result believes that she can walk independently when she cannot, she will fall on the floor when she tries. If, on the other hand, accurate self-assessment is preserved, she will use crutches or a wheelchair or get help from an attendant and so succeed in executing her intended task.

Note also that mental as well as physical capacities are subject to self-assessment and that among the mental capacities that agents may need to assess are the components of executive function itself. For example, if one is unable to detect errors well and at the same time is unable to assess one's own error-detection capacity, one will make multiple errors, possibly fatal ones, during task execution. If accurate self-assessment is preserved, however, and one knows that one has a problem with error detection, a number of strategies are available (such as double-checking one's work or seeking contemporaneous review by others). In general, then, preservation of accurate self-assessment allows for self-initiated use of compensatory strategies or prostheses (cognitive or physical) that result in successful task execution, whereas loss of accurate self-assessment precludes such action. One of the tragedies of many of the conditions that affect the anterior portions of the human brain is that the loss of the capacity for accurate self-assessment prevents the affected persons from adopting and using the very strategies that would compensate for their impaired functions, even when they possess all of the other physical and mental tools needed to use these strategies and use them well.

The capacity for self-assessment is commonly impaired by traumatic brain injury in which there is damage to both the frontal lobes and the inputs from other brain regions that are critical to normal frontal function. This is especially true of traumatic brain injury that results when an individual sustains a closed injury to the head of sufficient severity to cause significant acceleration-deceleration and/or torsion of brain structures as they move within the cranial vault in

response to the forces of injury. Common causes of injuries of this sort are car accidents and falls. The mild form of this injury is a concussion and affects primarily fibers (axons) ascending from the brainstem to the forebrain. The resulting syndrome includes deficits in arousal and attention/concentration, but much of executive function, including self-assessment, is usually well-preserved. As injury severity increases, however, the ascending fibers to frontal regions are particularly susceptible to further damage, and damage begins to accrue to frontal structures themselves through a variety of mechanisms that are beyond the scope of this paper. Damage to posterior structures may also occur, but this is often less severe than damage to frontal structures and may be absent altogether. Moreover, the prefrontal areas that mediate executive function are more susceptible than the primary motor areas in the more posterior portion of the frontal lobes.

As a consequence of their neuropathology, therefore, survivors of moderate or severe traumatic brain injury frequently show deficits in executive function, and these deficits are often out of proportion to their deficits in other neurologic functions. They may do quite well, for example, on many types of structured intelligence tests, walk freely with only mild problems in high-level balance and motor control, and have normal or near-normal perceptual function. However, their ability to develop, carry out, and monitor a rational, complex behavioral plan in an unstructured environment is often impaired, sometimes severely. Moreover, the behaviors encountered when this happens give us insight into their capacity to act as genuinely autonomous agents, as we believe our discussion of the following case will help to show.

THE CASE OF THE BARGE WORKER ("HENRY")

The case for consideration here is that of a 38-year-old man who sustained deep bifrontal brain contusions in a fall at work. "Henry," as we shall refer to him, worked as a laborer for a company that was involved in warehousing goods and shipping them down a river on a barge. His

responsibilities included stacking and securing heavy cartons in both the warehouse and on the barge. The accident occurred when he fell one storey from a platform in the warehouse onto the warehouse floor. He was initially comatose, and the admission CT scan of his brain revealed contusions in his caudate nuclei bilaterally. After emerging from coma and becoming medically stable, he was transferred to an inpatient brain injury rehabilitation program.

On admission to rehabilitation, Henry's thinking was slow and concrete. In addition, he showed the flattened affect, lack of spontaneity, and lack of initiation that are characteristic of persons with lesions in the area of his contusions. Also consistent with lesions in these areas, he showed impaired postural reflexes and a balance disorder that made him a significant falls risk, even when walking on a level surface. Moreover, he was unable to assess his ability to walk independently, although at the time his other cognitive deficits protected him from the potential adverse consequences of this, because he lacked the initiation to walk unaccompanied.

As he recovered and ability to initiate action returned to normal, Henry was fortunate in that he also experienced enough simultaneous improvement in balance and postural reflexes to walk safely on his own in a stable hospital (or home) environment. During the latter part of his hospital stay, he walked without supervision on the ward, and although his thinking remained somewhat slowed, he was able to direct himself independently to all of his scheduled therapies. Yet he continued to show residual high-level balance disorder to a degree that compromised his safety in physically unstable environments. In particular, he would not have been able to maintain his balance on a barge that was rocked by the wake of a passing boat.

Most important for our purposes, Henry's understanding of his impaired balance was strikingly limited in critical ways that affected his actions. When a balance failure was purposefully demonstrated to him in therapy, he was able to acknowledge verbally that he had lost his balance in that particular instance. However, when he was actually engaged in supervised physical

activity in the therapy gym, he never spontaneously took steps to stabilize himself in those situations in which he was at risk to lose his balance. Thus, the therapists constantly had to cue him to prevent him from falling while practicing motor tasks in unstable settings. Moreover, when asked about his overall ability to balance himself, he said that it was just fine, and neither the cuing nor his actual balance failures led him to modify that self-assessment. In other words, his global sense of himself as an agent did not incorporate an understanding of his balance disorder in any identifiable way.

Despite his deficits, however, both his motor function and his cognition recovered sufficiently to permit a return to work as a laborer under conditions in which it could be ensured that he would always be physically safe, even though his self-assessment deficit prevented him from grasping precisely what those conditions might be. For his part, returning to work was his primary goal, and he viewed working as an intrinsic and essential part of who he was as a person. When the time came to plan his return to work, his old company, which had regarded him as a reliable and loyal employee and which felt a sense of loyalty to him, developed a position for him as a "sweeper," cleaning the floor of the warehouse. In this position, he would not be exposed to any environment, such as heights or a rocking barge, that would be unsafe for a person with a balance disorder. However, although his primary goal was to return to full-time employment, he had a preference for his prior job, particularly on the barge, as he had always liked the water and the outdoors. His company would not rehire him in the old position, citing safety concerns. Because of his impaired self-assessment, however, Henry was unable to understand why the company would not let him have his old job back, and he refused to take the new position, leaving him unemployed.

This case illustrates several of the points we have been making.

At the level of specific actions, Henry's difficulty in walking on unstable surfaces illustrates how an agent's inaccurate capacity assessment can undermine his ability to act autonomously.

Again, the problem is not with the balance disorder itself but rather with Henry's inability to integrate the fact of his balance disorder into the generation of his actions. The deficit in accurate self-assessment creates a disconnect between what Henry sets out to do (namely, stay upright on a shifting surface) and what actually happens when he tries to do so. Inaccuracies in capacity self-assessment impair guidance control even at this concrete level.

At the level of the more global ability to lead one's life autonomously, this case illustrates how deficits in self-assessment can undermine the extent to which one's chosen course of action is one's own. Henry's persistent inability to integrate the relevance of his balance disorder into his decision-making kept him from being able to satisfy what he had identified as his most deeply held desire, which was to return to full-time work. Because of this disconnect, he kept getting in his own way.¹¹ Had he appreciated the relevance of his balance disorder, he would likely have taken the new job and satisfied his desire to be employed. But because of his breakdown in executive function, he was unable to comprehend why he would have to adjust his work expectations to achieve, as he could have, his more fundamental goal. It is this lack of accurate self-assessment—rather than his balance disorder *per se*—that kept Henry from being who he manifestly wanted to be, and thus from leading his own life.¹²

Finally, at a methodological level, the analysis of this case illustrates nicely how neurology and philosophy can mutually inform one another. For although the distinctions being drawn here are primarily conceptual, it was not until we discussed closely the empirical details of Henry's brain injury (as well as John's) that the distinction emerged between posterior brain deficits of awareness and anterior brain deficits of integration. Thus, as a matter of the "context of discovery," it was the examination of this neurologic distinction that led us to see that inaccurate self-assessment is not merely a matter of false beliefs but rather of a defective way of integrating knowledge about oneself into a global sense of oneself as the agent of particular actions. In addition, as

a matter of justification, there is a suggestive convergence of intuitions about what is essential to autonomy and what is essential to normal executive function, as revealed by the profound disruptions caused by breakdown of ordinary capacities for accurate self-assessment. As the set of neurocognitive capacities that gives coherence and structure to novel, prospective, goal-directed actions undertaken in the real world, executive function is essential if one is not to be restricted to overlearned behaviors in familiar settings or in settings structured by others. If autonomy entails genuine self-guidance, therefore, achieving full autonomy necessarily requires activation of executive function. That does not mean that every capacity that contributes to executive function is necessary for autonomy (or vice versa), but the foregoing discussion does suggest that there is some degree of isomorphism between autonomy and executive function, in that deficits in executive function seem to correspond directly and closely with loss of autonomy.

Once accurate self-assessment of capacity is seen as necessary for autonomy, it becomes clear why increasing it can enhance the autonomy of individuals without brain damage. In part, this is entailed already by acknowledging that we all have cognitive abilities that fall short of the ideal, that we all are, in some sense, impaired. To the extent to which we can avoid intending what we cannot do, our actions will be more a reflection of who we are and we will be more genuinely self-guiding.

OBJECTIONS

There are a variety of objections likely to be raised against the link we have drawn here between accurate self-assessment and autonomy. We consider three objections here.

OBJECTION #1: THAT THE REQUIREMENT OF ACCURATE SELF-ASSESSMENT IS UNREALISTICALLY DEMANDING

Many readers may find it unrealistic or overly demanding to expect that agents accurately assess their own abilities before they can count as

autonomous. Indeed, self-assessments are so often inaccurate that our view seems to have the problematic implication that too many clearly autonomous agents would seem no longer to count as autonomous.

This objection is misplaced in two regards. To begin with, it is based on an underestimation of the extent to which human action is routinely reflexive in the most mundane cases. One's self-assessment does not have to be explicit or conscious for it to be effective in the action-guiding role it plays. On the contrary, it typically operates in the background.¹³ Moreover, given how problematic inaccurate self-assessment is for even the most basic tasks in the cases of John and Henry, and given how many tasks require some accurate self-assessment for their successful completion, the fact that the great majority of us can navigate these tasks easily suggests that sufficiently accurate self-assessment must not be too hard to come by.

In addition, the fact that we fall short of perfectly accurate self-assessment does not show the requirement to be unrealistic. In this sense, the objection trades on blurring the role of autonomy as an ideal and as a practical threshold. As we are using the term, *autonomy* refers to a set of competencies that admit of degrees. For practical purposes, there will be cutoff points that specify when, for example, someone is competent to consent to medical treatment. Setting those thresholds too high would clearly be morally suspect. But there is nothing suspect about saying that a more accurate assessment of one's capacities would bring one closer to an ideal of optimal autonomy and that such approximation is valuable. Ultimately, of course, the value of pursuing such an ideal needs to be explained, as well as its value relative to other values (Haworth 1986, ch. 12), but we need not provide that account for the conception we are offering to be coherent.

OBJECTION #2: THAT THE REQUIREMENT OF ACCURATE SELF-ASSESSMENT IS BIASED

Even if accurate self-assessment is attainable for many, it still might be thought to be especially difficult for some, in a way that is biased and

exclusionary, thus violating fundamental liberal principles of inclusivity by privileging particularly reflective lifestyles. In particular, the fact that some people analyze themselves more than others should not provide any reason for thinking that they are more autonomous. Indeed, we should be particularly careful, especially in clinical settings, not to make the criteria for autonomy too lifestyle specific, given the way such approaches have been used to exclude and stigmatize marginal social groups or as a bias in favor of highly educated classes or subcultures.

This last point is well-taken, but we do not believe it represents an objection to our view. Note that, to reformulate the point made in response to the previous objection, inequalities in the level of autonomy attained do not necessarily entail inequalities in how people are treated. Insofar as we are concerned with legal or institutional authority, what matters is whether the autonomy of one's action is above the relevant threshold and whether one's capacity assessment is reasonable (rather than accurate), and nothing we say here rules out setting that threshold rather low.

But the objection also misses the mark by attributing to us a position that we do not hold, namely, that self-assessment is something that highly educated cultural elites do better than others. Note that, on our view, what is required is knowledge of one's capacities as they are relevant to the tasks one undertakes. Of course, if one is engaged in a life of philosophical contemplation, then the capacities one will need to assess are of an intellectual sort, and thus the assessment of them may be similarly intellectual. But homemaking is equally a domain in which being able to act as one intends requires accurate self-assessment. Indeed, given our emphasis on the mundane character of self-assessment, it seems that those who are actively engaged in the world will be at least as likely as those who engage in extensive and explicit reflection to have a good sense of their abilities.

There might, however, be contingent empirical connections between being able to lead one's life as one's own and engaging in certain practices of explicit reflection and the education that

enables such practices. But if that is the case—and it probably is, to some extent—then the conclusion to draw is surely that access (educational and otherwise) to those practices should be opened up, but not that the concept should be altered to render such individuals no longer deprived. Some people do, granted, make a lifestyle choice not to develop their capacities for self-assessment and, hence, for autonomy. And some cultures may choose to value autonomy less highly than others. But it is perfectly compatible with respect for pluralism of lifestyles and cultures to say that these choices make them less autonomous, as long as one does not assume—as we do not—that the value of autonomy trumps all other values.

OBJECTION #3: THAT INACCURATE SELF-ASSESSMENTS SOMETIMES PROMOTE AUTONOMY

Finally, it could be objected that accuracy in one's self-assessments is not as important to autonomy as we have been making it out to be, in that knowing the truth about one's weaknesses can sometimes be very debilitating. A number of psychological studies have suggested that self-aggrandizing illusions about one's abilities and performance may enhance one's feelings of self-confidence and sense of control over one's life, thereby enabling one to achieve more than one would have achieved if one had held a more accurate self-assessment (Taylor and Brown 1988). From this perspective, our insistence on accuracy may actually be counterproductive for autonomous agency, given the importance of a strong sense of optimism, self-esteem, or ambition for initiative, perseverance, and even performance.

There are three quick points we would like to make in reply. First, we have already noted that inaccuracies in self-assessment can be the result of underestimation as much as of overestimation. Cases in which self-doubt paralyzes a person do not count as counterexamples to our position if the self-doubt is unwarranted. Second, to the extent to which inflated estimations of one's ability to do something can be self-fulfilling, it seems implausible to say that the

self-assessment was inaccurate. There are metaphysical complexities here, but the burden of proof is on those who wish to show that upgrading one's capacity assessment provides a free-standing influence to the capacity in question. Alternatively, some account must be given of how it can be rational for an agent to hold a belief (about her ability to do something) that she does not think will come to be true.¹⁴ And finally, there are reasons to be cautious about the literature on self-esteem and the "power of positive thinking." For however much it may be true that the global self-doubt of depressed individuals can impair their ability to initiate activity and however much subjective happiness correlates with positive self-esteem, it is unclear that inflated self-assessments yield benefits in contexts where what is required is not just effort but competence. With regard to numerous tasks such as driving a car or investing in the stock market, overconfidence actually tends to undermine one's ability to achieve one's goals.¹⁵

Of course, there are still complex issues regarding how one judges the accuracy of someone's self-assessment (or one's own). Some cases will be much more straightforward than others. For example, in the case of John, we know both that his optic nerves are severed and that there is damage to regions of the brain that are required for self-assessment. This knowledge allows us to attach a high degree of confidence to our external judgment that John's self-assessment is inaccurate. Other cases are much harder to determine, such as a 10-year-old's sense that he can become President of the United States. Here, our judgment of the accuracy of the self-assessment will vary with the quality of the evidence: How complicated is the causal chain that would lead to the result? How good is the individual's assessment of his own abilities more generally, in other contexts? Is it possible that the agent's pursuit of an unattainable ideal reflects not a confused self-assessment but a clearheaded commitment to the intrinsic value of the pursuit itself? Evaluating these factors and forming a judgment is difficult and a certain degree of indeterminacy may well be unavoidable, but that does not undermine the usefulness of the concept

of accurate self-assessment. And as a practical matter, the natural thing to say is that the greater the degree of uncertainty or complexity regarding external judgments of the accuracy of a person's self-assessment, the more caution is called for in withholding the attribution of autonomy to that person.

CONCLUSION

We have argued here for including accurate self-assessment among the necessary conditions for personal autonomy. Unless one can evaluate and appreciate the extent to which one has the capacities required for an intended course of action, one's pursuit of that course of action lacks the self-guiding character distinctive of fully autonomous action. As we have emphasized throughout, this capacity for accurate self-assessment is a matter of degree. By analyzing, at one end of this continuum, clinical cases of traumatic brain injury, we have underscored how serious damage to the neuropsychological substrate for executive function (and for self-assessment in particular) makes autonomous action impossible in certain regards. As is often the case, however, analyzing brain damage also reveals much about normal function. And part of our point has been to demonstrate the significant extent to which those of us without brain damage ordinarily rely on a sense of our own capacities in guiding our actions. Typically, we do know our own strength, and this enables us to use whatever strength we have to accomplish our intended tasks. Our self-assessments are sometimes off-the-mark, but to the extent to which we can reduce the inaccuracies in our self-assessments, we are better able to guide our actions in a fully autonomous sense.

In making our empirically grounded conceptual points, we have set to one side the important but distinct task of proposing guidelines for when an agent should or should not count as autonomous for various legal, bioethical, or other purposes. But our argument here does have at least one important implication for those more practically oriented discussions. For if we are right about the crucial importance of accurate self-

assessment to autonomous action, no practical criteria for autonomy can be adequate without some consideration of individuals' capacities for understanding themselves and integrating that self-understanding into their intentions for action. In other words, the model of the competent, autonomous agent must be that of a reflexive agent.

NOTES

1. Although the names have been changed to preserve anonymity, the cases discussed in this paper are actual cases from Warren Lux's clinical neurologic practice.

2. For an extensive discussion of the full range of disturbances in awareness of deficit due to brain disease, including the anosognosia associated with parietal syndromes as well as the more global self-awareness disorders associated with frontal syndromes, see Prigatano and Schacter (1991), especially the chapters by Stuss (1991) and Kihlstrom and Tobias (1991). The self-assessment problems with which we are concerned in this paper are part of the spectrum of frontal self-awareness disorders.

3. The distinction here is not simply analytic but actually has a neurologic correlate. Sensory perceptual disturbances due to parietal, particularly right parietal, disease may involve an inability to discern task parameters accurately, and thus the resulting errors in self-assessment would have a different origin than those caused by the kinds of brain injuries with which we are concerned here. In our view, however, autonomy would be impaired in either case, albeit differently and perhaps to a different degree.

4. This idea of the practical reflexivity of human agency is defended across a variety of schools of thought, although it is probably most prominent in broadly Kantian (and Hegelian) approaches to agency (see, e.g., Brandom 1994; Habermas 1984; Honneth 1995; Korsgaard 1996; and Velleman 1989, 2000, ch. 1, 6, 8). But the same idea underlies Fischer and Ravizza's talk of taking responsibility for the mechanism from which one's actions flow (1998, ch. 8).

5. See Christman (2001), Dworkin (1988), Frankfurt (1988), Mele (1995), and Meyers (1989). A recent overview can be found in the

editor's Introduction to Christman and Anderson (forthcoming).

6. Terence Penelhum highlights with particular sharpness the difficulty with the language of desires being "one's own": "To say that the desire is not one's own and mean this literally is to say something obviously false: for the desire is operative and therefore exists, and is not someone else's" (1971, 670).

7. There are actually very interesting connections between inaccurate self-assessment of one's capacities and difficulties in assessing performance, especially of detecting errors in one's behavior. When individuals are unable to recognize when their own performance departs from their intended actions, they are cut off from direct evidence for their incapacity. It is possible, of course, to develop an accurate, if somewhat artificial, sense of one's capacity while still being unable to assess one's performance oneself, but to do this one must rely on others' reports of one's performance and also on higher order cognitive processes. In another paper, we intend to take up this phenomenon, drawing on the case of a prefrontally brain-injured individual who was unable to monitor the volume, tone, and manner of his own speech but knew this about himself and deliberately undertook compensatory strategies. This sort of case is actually quite rare, however.

8. Penelhum (1971), for example, claims that an agent representing a desire as "external to himself" is engaging in a form of "moral trickery."

9. See Millgram (1997) on the inferential nature of "genuine desires." We do, of course, have action-guiding mental states that are not well-embedded inferentially: whims, for example, or a taste for pistachio ice cream. But it is enough to say, then, that inaccurate self-assessment demotes one's desires to the status of "whim." For a further discussion of the role of integration in autonomy, see Anderson (2003).

10. For discussion of executive function and the capacities that comprise it, see Lezak (1989, ch. 7).

11. The idea that autonomy is threatened centrally by cases of one getting in one's own way—

of being volitionally self-defeating—is perhaps the dominant theme in Harry Frankfurt's recent work (see esp. 1988, ch. 12; and 1999, ch. 8, 9, 11, 14).

12. Of course, given the complexity and systematicity of human motivation, it is impossible to predict for certain that Henry would have accepted the new job if he were able to appreciate fully his balance disorder. He did, for example, present some of the general cognitive rigidity that often accompanies traumatic brain injury. But the primary source of his anger at the company was not that the new job was beneath his dignity, but rather that he could not see any reason why they would not let him return to working on the barge, a confusion traceable to his inaccurate capacity assessment.

13. For a particularly insightful analysis of the way in which agents guide their actions and utterances in light of a reflexive sense of their position in a social setting, see Garfinkel (1967).

14. The materials for constructing a strong argument for why self-fulfilling beliefs should not generally be considered cases of inaccurate self-assessments can be found in Velleman's discussion of "reflective expectations" (1989, esp. 64). See also Williams (1973) and Millgram (1997, esp. 32–35).

15. One particularly striking analysis of the dangers of overestimating one's abilities can be found in the work of behavioral economist Terrence Odean. He has shown that individual stock market investors tend to be overconfident (especially if they are men) about the quality of the information they have about the companies whose stock they trade, and that this has disastrous consequences for their investments (Odean 1998). He draws here on the work of others that shows a similar phenomenon of overconfidence regarding driving ability. Note that even in their influential article on the productiveness of illusions, Taylor and Brown (1988) admit that "The evidence for direct effects of positive illusions on intellectual functioning is sparse" (198). The perils of overestimation should not, however, be understood as a challenge to recognizing the importance of baseline positive relations to self, as Trudy Govier (1993) has argued for self-trust

and Axel Honneth (1995) has argued for self-respect, self-esteem, and basic self-confidence. (See also Anderson and Honneth forthcoming.)

ACKNOWLEDGMENT

Earlier versions of this paper were presented at Washington University's Philosophy-Neuroscience-Psychology program, the Center for Clinical Bioethics at Georgetown University, the 2002 APA Pacific Division Meetings, the Washington University Disabilities Studies Group, and the Utrecht University Philosophy Department. We have benefited greatly from comments from those audiences and others. We especially thank Larry May, Pauline Kleingeld, Andy Clark, Brian Keeley, Ilya Farber, Daniel Weiskopf, John Christman, Dominic Murphy, James Harold, George Graham, Roberto Dell'Oro, Philip Robbins, Walton Schalick, Robert Pollack, Stanley Finger, Shaun Nichols, and Alan Lambert.

REFERENCES

- Anscombe, G. E. M. 1985. *Intention*. 2nd ed. Ithaca, NY: Cornell University Press.
- Anderson, J. 2003. Autonomy and the authority of personal commitments: From internal coherence to social normativity. *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action* 6:90–108.
- Anderson, J., and A. Honneth. Forthcoming. Autonomy, vulnerability, recognition, and justice." In *Autonomy and the challenges to liberalism: New essays*, ed. J. Christman and J. Anderson. New York: Cambridge University Press.
- Berofsky, B. 1995. *Liberation from self: A theory of personal autonomy*. New York: Cambridge University Press.
- Brandom, Robert B. 1994. *Making it explicit: Reasoning, representing, and discursive commitment*. Cambridge, MA: Harvard University Press.
- Christman, J. 1991. Autonomy and personal history. *Canadian Journal of Philosophy* 20:1–24.
- . 2001. Liberty, autonomy, and self-transformation. *Social Theory and Practice* 27:185–206.
- Christman, J., and J. Anderson., eds. Forthcoming. *Autonomy and the challenges to liberalism: New essays*. New York: Cambridge University Press.
- Clark, A., and D. Chalmers. 1998. The extended mind. *Analysis* 58:7–19.
- Dworkin, G. 1988. *The theory and practice of autonomy*. New York: Cambridge University Press.
- Fischer, J. M., and M. Ravizza. 1998. *Responsibility and control: A theory of moral responsibility*. New York: Cambridge University Press.
- Frankfurt, H.G. 1988. *The Importance of what we care about: Philosophical essays*. New York: Cambridge University Press.
- . 1999. *Necessity, volition, and love*. New York: Cambridge University Press.
- Garfinkel, H. 1967. *Studies in ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall.
- Govier, T. 1993. Self-trust, autonomy, and self-esteem. *Hypatia* 8:99–120.
- Habermas, J. 1984. *The theory of communicative action*, vol. 1, trans. T. McCarthy. Boston: Beacon Press.
- Haworth, L. 1986. *Autonomy: An essay in philosophical psychology and ethics*. New Haven, Conn: Yale University Press.
- Honneth, A. 1995. *The Struggle for recognition: The moral grammar of social conflicts*, trans. J. Anderson. Cambridge, UK: Polity Press.
- Kant, I. 1785/1997. *Groundwork of the metaphysics of morals*, trans. M. Gregor. Cambridge, UK: Cambridge University Press. [German original 1785.]
- Kihlstrom, J. F., and B. A. Tobias. 1991. Anosognosia, consciousness, and the self. In *Awareness of deficit after brain injury: Clinical and theoretical issues*, ed. G. P. Prigatano, and D. L. Schacter, 198–222. New York: Oxford University Press.
- Korsgaard, C. 1996. *The Sources of normativity*. Cambridge, UK: Cambridge University Press.
- Lezak, M. D. 1989. *Assessment of the behavioral consequences of head trauma*. New York: Alan R. Liss.
- Mele, A. 1995. *Autonomous agents: From self-control to autonomy*. New York: Oxford University Press.
- Meyers, D. T. 1989. *Self, society and personal choice*. New York: Columbia University Press.
- Millgram, E. 1997. *Practical induction*. Cambridge, Mass: Harvard University Press.
- Odean, T. 1998. Volume, volatility, price, and profit: When all traders are above average. *Journal of Finance* 53:1887–1934.
- Penelhum, T. 1971. The importance of self-identity. *Journal of Philosophy* 63:667–679.
- Prigatano, G. P., and D. L. Schacter. 1991. *Awareness of deficit after brain injury: Clinical and Theoretical issues*. New York: Oxford University Press.
- Rovane, C. 1998. *The Bounds of agency: An essay in revisionary metaphysics*. Princeton, NJ: Princeton University Press.
- Stuss, D. T. 1991. Disturbance of self-awareness after frontal lobe damage. In *Awareness of deficit after brain injury: Clinical and theoretical issues*, ed. G.

- P. Prigatano, and D. L. Schacter, 63–83. New York: Oxford University Press.
- Taylor, S. E., and J. D. Brown. 1988. Illusion and well-being: a social psychological perspective on mental health. *Psychological Bulletin* 103:193–210.
- Velleman, J. D. 1989. *Practical reflection*. Princeton, NJ: Princeton University Press.
- Velleman 2000. *The possibility of practical reason*. Oxford, UK: Oxford University Press.
- Williams, B. 1973. Deciding to believe. In *Problems of the self*, 136–151. Cambridge: Cambridge University Press.
- Wolf, S. 1990. *Freedom within reason*. New York: Oxford University Press.