Consciousness and Free Will: A Critique of the Argument from Introspection

Gregg Caruso
Corning Community College

In the free will debate, libertarians have put a great deal of emphasis on our *conscious feeling of freedom* and our introspective abilities. In fact, many libertarians have suggested that our introspection of the decision-making process, along with our strong feeling of freedom, provides some kind of evidence for the existence of free will. As Ledger Wood describes the libertarian argument:

> Most advocates of the free will doctrine believe that the mind is directly aware of its freedom in the very act of making a decision, and thus that freedom is an immediate datum of our introspective awareness. "I feel myself free, *therefore*, I am free," runs the simplest and perhaps the most compelling of the arguments for freedom (1941: 387).

We can call this argument the *introspective argument for free will*. The introspective argument essentially maintains that, upon introspection, we do not *seem* to be causally determined—instead, we *feel* that our actions and decisions are freely decided by us—hence, we *must* be free. Libertarians, especially agent-causal theorists, take this introspective datum as their main evidence in support of free will.[1]

This kind of argument only works, however, if we assume the data is veridical. But, as we'll see, there is reason to doubt the reliability of such introspective evidence. How do we know that our feeling of freedom isn't an illusion? How do we know that what we introspect is accurate? Libertarians seldom question the "I feel myself free, therefore, I am free" argument. In fact, the introspective argument can be found throughout the literature. Given this, it is important that we investigate both the role of consciousness and the accuracy of introspection. Perhaps a closer examination of these issues will reveal that the nature of consciousness, rather than supporting free will further impugns it.

I. Consciousness and Free Will

Although libertarians put a great deal of emphasis on consciousness when it comes to introspecting our own freedom, they ironically overlook the importance of consciousness when it comes to explaining its role in *producing* free actions. Timothy O'Connor,

for example, seems to be aware of this shortcoming when he writes:

> Something the philosopher ought to be able to provide some general light on is how consciousness figures into the equation. It is a remarkable feature of most accounts of free will that they give no essential role to conscious awareness. One has the impression that an intelligent automata could conceivably satisfy the conditions set by these accounts—something that is very counterintuitive (2000: 122).

I share O'Connor's surprise at the fact that consciousness has not played a larger role in accounts of free will, especially given the obvious importance of conscious awareness. It truly is counterintuitive to think that one could exercise free will while at the same time being wholly unaware that they are doing so. An intelligent automaton *cannot* and *should not* be the paradigm of a free agent. One could even argue that it is logically inconceivable to imagine an *automaton*—a creature that lacks all conscious awareness—that has freedom.[2]

If libertarian accounts of freedom wish to be successful, then, they *need* to show that *one* of the functions of consciousness is that it somehow exercises free will.[3] Given the requirement of conscious awareness, it is a sad state of affairs when libertarians like O'Connor dedicate no more than a few lines to the issue. O'Connor himself only presents one, very vague proposal. He claims:

> The agency theorist can conjecture that *a* function of biological consciousness, in its specifically human (and probably certain other mammalian) manifestations, is to subserve the very agent-causal capacity I sketched in previous chapters (2000: 122).

Beyond this, O'Connor doesn't explain *how* or *in what way* consciousness *subserves* agent-causal powers. And after an exhaustive examination of the literature, I have been unable to find any substantial account of the role consciousness plays in libertarian freedom.

I can only conclude, then, that libertarianism lacks a complete story. On the one hand they appeal to our conscious feeling of freedom as evidence of free will, while on the other hand they neglect to explain the role and importance of consciousness. O'Connor's comments simply amount to the following two claims: (1) That the "self-determining capacity [required for libertarian freedom] strictly requires conscious awareness" (2000: 122); and (2) that *somehow* consciousness aids in this capacity. I point this out only to expose one (often overlooked) problem with libertarianism—they have provided us with nothing but a promissory note when it comes to the crucial question of explaining the role of consciousness in the exercise of agent-causal powers.

Returning, however, to the introspective argument, from the fact that

I *feel* myself free, it does not necessarily follow that I *am* free. The feeling could be an illusion. It is a fallacy to think that one could establish a metaphysical conclusion from phenomenology alone. More than an appeal to our introspective experience would be needed to prove that we actually *enjoy* agent-causation. That said, what this argument *does* show is that people often infer their own freedom from their introspective phenomenology of freedom. Why is this so? Well one possible answer is that people implicitly believe that we have access to all the causal factors and the causal processes underlying our own decision-making. If people were to believe in such introspective transparency, then it would be appropriate, given the above phenomenology, for them to infer that they are undetermined. For if one introspects no deterministic processes underlying one's decision making, and one also thinks that if there *were* a deterministic process one *would* introspect it, one could infer that there is no deterministic process. This is, I believe, part of the story of why we feel free.[4] It is important then to examine whether we really are transparently and infallibly aware of all our mental activity, for it is our belief in the transparency and infallibility of consciousness that gives the introspective argument whatever power it possesses.

Given that the majority of support for free will comes from our introspective awareness of the decision-making process, and consciousness appears to be a necessary condition for free will, it would seem that an account of the mind which claims that all mental states are conscious states would be more accommodating to defenders of free will. Such an account of the mind can be traced back to the father of modern libertarianism, René Descartes. One can find at the heart of Descartes' philosophy of mind three main theses:

(1) That the mind and body are two mutually exclusive, interacting, non-identical substances—the mind being a completely nonphysical substance.

(2) That there is nothing in our mind of which we are not conscious; i.e., all mental states are conscious states.

(3) Our knowledge of our own mental states is certain and infallible; our judgments about them cannot be erroneous.

The first of these theses, of course, has been widely criticized. Despite a retreat from the metaphysics of substance dualism, however, the rest of the Cartesian concept of mind remains largely intact when it comes to theorizing about consciousness and free will. Theses (2) and (3) combined

amount to the claim that all mental states are conscious and that such consciousness is infallible. Essentially this is the belief in the *transparency* and *infallibility* of consciousness.[5] From a first person point of view these two theses seem to make sense. We are conscious of our mental states in a way that seems, at least subjectively, to be direct, immediate, and infallible.

What we need to investigate here then is whether our phenomenology, which seems to support these two theses, is accurate or not. I will argue that it is not. From a first-person point of view, it may *seem* as though we are aware of all our mental states—including mental reasoning and decision making—in an immediate, direct, and infallible way, but from a third-person point of view we can often see that this is not the case. I believe that it is partly because consciousness appears transparent from the first-person point of view that we impart so much power to the conscious will. If it turns out, however, that consciousness is *not* transparent then the introspective argument for free will would lose its force.

II. Mentality and Consciousness

Let us begin with the claim that all mental states are conscious states. This conception of mentality and consciousness not only claims that consciousness is an essential property of mental states, but also that consciousness is the mark of the mental. For on the Cartesian concept of mind, what makes a state a mental state is its being a conscious state. States that are not conscious are also not mental. This, however, has significant theoretical drawbacks. Some, like David Rosenthal (2005), have argued that if consciousness is what makes a state a mental state, consciousness will not only be an intrinsic, nonrelational property of all mental states, it will also be impossible to analyze. Since the Cartesian concept of mind tacitly conflates mentality and consciousness—thereby making consciousness essential to all mental states—no reductive explanation of consciousness can be given in terms of other higher-level cognitive or mental processes (see Rosenthal, 1986, 2002).

If this is true, then any theory of consciousness that equates mind and consciousness faces a serious explanatory problem. But what is equally troubling (or perhaps even more troubling) is what accepting this equivalence means for understanding the mind itself, not only consciousness. If we were to equate mind and consciousness, we would then have to understand mental processes in terms of consciousness. "So accepting the equivalence would also prevent us from ever developing an informative account of mind" (Rosenthal, 2002: 237). We would be unable to investigate mental processes without at the same time investigating conscious processes. This, I believe, is not only theoretically unacceptable it is also empirically unjustifiable!

There is more than ample reason to believe that not *all* mental states are conscious states. Many types of mental states—such as thoughts, desires, beliefs, judgments, and intentions—often occur without being conscious. Both common sense and cognitive science typically posits mental states that are not conscious to explain certain behaviors and cognitive capacities. The most widely accepted of these unconscious mental states are intentional states. There are not only clinical and experimental results which provide good reason to hold that beliefs and desires exist that are not conscious, but everyday folk psychology makes much use of intentional states that are not conscious to explain the actions of others.[6] One could even argue that the majority of our intentional states are probably nonconscious.[7]

The work of Timothy Wilson, John Bargh, Ap Dijksterhuis, Benjamin Libet, and others,[8] has shown that "the same higher mental processes that have traditionally served as quintessential examples of choice and free will—such as goal pursuits, judgment, and interpersonal behavior—have been shown recently to occur in the absence of conscious choice or guidance" (Bargh & Ferguson, 2000: 926). It is no longer believed that *only* lower-level processing can occur outside the reach of consciousness. There is now growing evidence that a great deal of higher-level mental functioning is also nonconscious. Psychologists and cognitive scientists have accumulated a great deal of evidence for determinism by demonstrating that high-level mental and behavioral processes "can proceed without the intervention of conscious deliberation and choice" (Bargh & Ferguson, 2000: 925). All of this mounting research, I believe, indicates that high level unconscious cognitive states—states that are best described as *mental*—actively, and frequently, play a role in human behavior.[9] The Cartesian thesis, then, that all mental states are conscious states—though perhaps supported by phenomenology—has multiple strikes against it and is probably false.[10]

III. Knowing Thy Self: Consciousness and Self-Reports

In addition to the assumption that all mental states are conscious, libertarian and folk psychological accounts of consciousness usually make the related assumption that consciousness provides us with infallible knowledge of our (conscious) mental states. The claim of infallibility is another part of the traditional Cartesian concept of mind.[11] From a first-person point of view, this assumption seems to makes sense. Who else, we feel, is in a better position to know which mental states we are in than ourselves?

It is often assumed, almost at a definitional level, that we are immediately, directly, and infallibly connected to the content of our own minds. From a first-person point of view, it never seems as though consciousness

and mentality come apart. Subjectively, it never seems to us that consciousness mischaracterizes or misidentifies the mental states we are in. It is hard for us to believe that our consciousness can mislead us about the nature of our own minds or that we can be in mental states that we are unaware of.

Although this is undoubtedly how things *seem* from a first-person point of view, I do not think we can rely on phenomenology alone at the exclusion of all other information. There is a great deal of evidence suggesting that we are not always the best judges of what's going on in our own minds.[12] Researchers, for example, are increasingly realizing that the mental states and processes that they are interested in measuring are not always consciously accessible to their participants, forcing them to rely on alternative methods.[13] The introspective method—i.e., the method of relying on the introspective reports of subjects—has, in fact, come under attack numerous times throughout the history of psychology (See, e.g., Nisbett & Wilson, 1977; Lieberman, 1979; Jack & Roepstorff, 2002). In attitudes research, for example, a number of researchers now argue that people could simultaneously possess different implicit and explicit attitudes towards the same object—with self-reports measuring only the explicit attitude (e.g., Wilson, Lindsey & Schooler, 2000). This has led some to develop implicit measures to explore the nature of these attitudes and people's awareness of them (see, e.g., Greenwald, McGhee, & Schwartz, 1998).[14]

Our ability to know our own mental states is limited and fallible. People have access to many of their mental states, no doubt, but there is also a pervasive *adaptive unconscious* that is often inaccessible via introspection (Wilson, 2002, 2003). In addition, consciousness, which is accessible to introspective reports, does not always represent our mental states and processes accurately. There is good reason to believe that individuals often confabulate stories for why they do certain things. When this happens, one's first-person reports fail to match the actual causes for their action. This has been shown to happen, for example, in hypnotized subjects. After being hypnotized, subjects can enact a posthypnotic suggestion—e.g., "when you awake you will immediately crawl around on your hands and knees." When asked what they are doing, subjects almost immediately generate a rationale—"I think I lost an earring down here" (Gazzaniga, 1985; Hilgard, 1965; Estabrooks, 1943). From a first-person point of view, these individuals are conscious of a particular reason for why they are doing what they are doing, but from a third-person point of view we can see that this is not the real cause of their action. Similar examples of confabulation have also been found in "split brain" patients (Gazzaniga & LeDoux, 1978) and patients with Korsakoff's syndrome—a form of organic amnesia where people

lose their ability to form memories of new experiences (Sacks, 1987).[15]

Although it may be tempting to think such confabulation is limited to these rare occasions, some theorists have suggested that similar confabulation occurs throughout everyday life (see, e.g., Nisbett & Wilson, 1977; Gazzaniga & LeDoux, 1978; Wilson, 2002). These theorists argue that our conscious selves often do not fully know why we do what we do and thus have to confabulate stories and create explanations. In one of the most famous papers on the subject, Nisbett and Wilson (1977) placed subjects in identical situations save for the fact that one or two key features were varied. They observed that although these key features influenced people's judgments or behavior, when asked to explain why they responded the way they did, subjects remained unaware of the varied features and instead confabulated different explanations for their behavior. Studies like this seriously question the accuracy of consciousness awareness, because they reveal that consciousness does *not* provide us with transparent and infallible knowledge of our own minds. We may consciously think we are doing something because of reason *X*, when in reality we are doing it because of reason *Y*.[16]

IV. Conclusion

What does this mean for the introspective argument for free will? I believe it shows that we cannot rely on our conscious experience *alone* to determine what the true causes of our actions are. We are often unaware of important causal determinates. The fact that we do not *feel* causally determined, or that we are not consciously aware of the various (internal and external) influences on our behavior, does not mean such determinates do not exist. Worse still, if consciousness can confabulate and/or misrepresent the causes for our choices and/or actions, then to rely on such conscious data to infer our own freedom would be a mistake. Whatever persuasiveness the introspective argument originally had depended on the assumption that we had direct, infallible access to our own decision-making process. The argument assumes that consciousness reveals everything about our mental functioning, or at least everything relevant to the issue at hand. If what I've argued here is correct, this is not the case. What we are conscious of—and hence, what we can report on—is not always in line with what is otherwise going on mentally. If it turns out, as recent research suggests, that a great deal of mental activity is controlled by unconscious mental states, and that we can also misrepresent and confabulate the states we are in, the libertarian argument from introspection loses all its force.

Works Cited

Aarts, H., and Ap Dijksterhuis (2000). "Habits as Knowledge Structures: Automaticity in Goal-Directed Behavior," *Journal of Personality and Social Psychology*, vol. 78: 53-63.

----- (2003). "The Silence of the Library: Environment, Situational Norm, and Social Behavior," *Journal of Personality and Social Psychology* 84, 1: 18-28.

Caruso, Gregg (2005). "Sensory States, Consciousness, and the Cartesian Assumption," in *Descartes and Cartesianism*, eds. Nathan Smith and Jason Taylor, Cambridge Scholars Press: 177-199.

----- (in progress). *Free will and Conscious Illusion: A Determinist Account of the Cognitive Illusion of Free Will*.

Descartes, René (1964-75). *Oeuvres de Descartes*, eds. Charles Adam and Paul Tannery, revised edition, Paris: Vrin/ C.N.R.S.

----- (1984-85). *The Philosophical Writings of Descartes Vols. 1-3*. Trans. John Cottingham, Robert Stoothoff, and Dugald Murdoch. Cambridge: Cambridge University Press.

Bargh, John. (1989). "Conditional Automaticity: Varieties of Automatic Influences in Social Perception and Cognition." In J.S. Uleman & J.A. Bargh (Eds.), *Unintended Thought*, New York: Guilford Press: 3-51.

----- (1990). "Auto-motives: Preconscious Determinants of Social Interaction." In E.T. Higgins & R.M. Sorrentino (Eds.), *Handbook of Motivation and Cognition* (Vol. 2), New York: Guilford Press: 93-130.

----- (1994). "The Four Horsemen of Automaticity: Awareness, Efficiency, Intention, and Control in Social Cognition." In R.S. Wyer, Jr., & T.K. Srull (Eds.) *Handbook of Social Coignition* (2nd ed.), Hillsdale, NJ: Erlbaum: 1-40.

----- (1996). "Automaticity in Social Psychology." In E.T. Higgins & A.W. Kruglanski (Eds.), *Social Psychology: Handbook of Basic Principles*, New York: Guilford Press: 169-183.

----- (1997). "The Automaticity of Everyday Life." In R.S. Wyer,

Jr. (Ed.), *The Automaticity of Everyday Life: Advances in Social Cognition*, Mahwah, NJ: Erlbaum: 1-61.

Bargh, J., and T.L. Chartrand (1999). "The Unbearable Automaticity of Being," *American Psychologist* 54, 7: 462-479.

Bargh, J., and M. Ferguson (2000). "Beyond Behaviorism: On the Automaticity of Higher Mental Processes," *Psychological Bulletin* 126, 6: 925-945.

Bargh, J., P.M. Gollwitzer, A. Lee-Chai, K. Barndollar, & R. Trotschel (2001). "The Automated Will: Nonconscious Activation and Pursuit if Behavioral Goals," *Journal of Personality and Social Psychology*, 81: 1014-1027.

Bechara, A., H. Damasio, D. Tranel, and A.R. Damasio (1997). "Deciding Advantageously Before Knowing Advantageous Strategy." *Science* 275: 1293-95.

Chartrand, T.L., and J. Bargh (1996). "Automatic Activation of Impression Formation and Memorization Goals: Nonconscious Goal Priming Reproduces Effects of Explicit Task Instructions," *Journal of Personality and Social Psychology* 71: 464-478.

Dijksterhuis, Ap (2004). "I Like Myself But I Don't Know Why: Enhancing Implicit Self-Esteem by Subliminal Evaluative Conditioning," *Journal of Personality and Social Psychology* 86, 2: 345-355.

Dijksterhuis, Ap, and Ad van Knippenberg (1998). "The Relation Between Perception and Behavior, or How to Win a Game of Trivial Pursuit," *Journal of Personality and Social Psychology* 74, 4: 865-877.

Dijksterhuis, Ap, R. Spears, T. Postmes, D.A. Stapel, W. Koomen, A. van Knippenberg, and D. Scheepers (1998). "Seeing One Thing and Doing Another: Contrast Effects in Automatic Behavior," *Journal of Personality and Social Psychology* 75: 862-871.

Ericsson, K.A., and H.A. Simon (1980). "Verbal Reports as Data," *Psychological Review* 87: 215-251.

Estabrooks, G.H. (1943). *Hypnotism*. New York: E.P. Dutton.

Gavanski, I., and C. Hoffman (1987). "Awareness of Influences on One's Own Judgments: The Role of Covariation Detection and Attention to the Judgment Process," *Journal of Personality and Social Psychology* 52: 453-463.

Gazzaniga, M.S. (1985). *The Social Brain*. New York: Free Press.

Gazzaniga, M.S., and J.E. LeDoux (1978). *The Integrated Mind*. New York: Plenum.

Greenwald, A.G., D.E. McGhee, and J.L.K. Schwartz (1998). "Measuring Individual Differences in Implicit Cognition: The Implicit Association Test," *Journal of Personality and Social Psychology* 74, 6: 1464-1480.

Hilgard, E.R. (1965). *Hypnotic Susceptibility*. New York: Harcourt, Brace & World.

Jack, A.I., and A. Roepstorff (2002). "Introspection and Cognitive Brain Mapping: From Stimulus-Response to Script-Report," *Trends in Cognitive Science* 6, 8: 333-339.

LeDoux, J. (1996). *The Emotional Brain: The Mysterious Underpinnings of Emotional Life.* New York: Simon and Schuster.

Libet, Benjamin. (1985). "Unconscious Cerebral Initiative and the Role of Conscious Will in Voluntary Action," *Behavioral and Brain Science* 8: 529-66.

----- (1987). "Are the Mental Experiences of Will and Self-Control Significant for Performance of a Voluntary Act?," *Behavioral and Brain Science* 10: 783-91.

----- (1993). "The Neural Time Factor in Consciousness and Unconscious Events," in *Experimental and Theoretical Studies of Consciousness* (Ciba Foundation Symposium; 174). London: John Wiley and Sons.

Lieberman, David A. (1979). "Behaviorism and the Mind: A (Limited) Call for a Return to Introspection," *American Psychologist* 34: 319-333.

Nisbett, R.E., and T.D. Wilson. (1977). "Telling More Than We Can Know: Verbal Reports on Mental Processes," *Psychological Review* 84: 231-259.

O'Connor, Timothy. (1995). "Agent Causation," in *Agents, Causes*

and Events: Essays on Free Will and Indeterminism, Timothy O'Connor (ed.). Oxford: Oxford University Press: 173-200.

----- (2000). *Persons and Causes: The Metaphysics of Free Will*. New York: Oxford University Press.

Rosenthal, David M. (1986). "Two concepts of consciousness," *Philosophical Studies* 49, 3 (May): 329-359; reprinted in *The Nature of Mind*, (ed.) David Rosenthal. New York: Oxford University Press, 1991: 462-477.

----- (2002). "Consciousness and Mind," *The Jerusalem Philosophical Quarterly* 51 (July): 227-251.

----- (2004). "Varieties of Higher-order Theory," in *Higher-Order Theories of Consciousness*, (ed.) Rocco Gennaro. Amsterdam: John Benjamins.

----- (2005). *Consciousness and Mind*. New York: Oxford University Press.

Sacks, Oliver (1987). *The Man Who Mistook His Wife For a Hat and Other Clinical Tales*. New York: Harper and Row.

Searle, John. (1990). "Consciousness, Explanatory Inversion, and Cognitive Science," *Behavioral and Brain Sciences* 13: 585-642.

----- (1992). *The Rediscovery of the Mind*. Cambridge: MIT Press/ Bradford Book.

Smith, E.R., and F.D. Miller (1978). "Limits on Perception of Cognitive Processes: A Reply to Nisbett and Wilson," *Psychological Review*, 85: 355-382.

Steele, Claude, and Joshua Aronson (1995). "Stereotype Threat and the Intellectual Test Performance of African Americans," *Journal of Personality and Social Psychology* 69, 5: 797-811.

Taylor, Richard. (1992). *Metaphysics*, 4th edition. Englewood Cliffs, NJ: Prentice-Hall.

Wilson, Timothy (2002). *Strangers to Ourselves: Discovering the Adaptive Unconscious*. Cambridge, Mass.: The Belknap Press of Harvard University Press.

----- (2003). "Knowing When to Ask: Introspection and the Adaptive Unconscious," *Journal of Consciousness Studies* 10: 131-140.

Wilson, T., S. Lindsey, and T.Y. Schooler (2000). "A Model of Dual Attitudes," *Psychological Review*, 107: 101-126.

Wilson, T., and J.I. Stone (1985). "Limitations of Self-knowledge: More on Telling More Than We Can Know." In P. Shaver (ed.), *Review of Personality and Social Psychology Vol.6*. Beverly Hills: Sage: 167-183.

Wood, Ledger (1941). "Determinism: Free Will Is an Illusion." *Philosophy* 16: 386-89.

Notes

[1] Agent-causal theorists traditionally take our feeling of freedom and self-exertion as evidence that we are agent-causes. Timothy O'Connor, for example, writes: "[T]he agency theory is appealing because it captures the way we experience our own activity. It does not seem to me (at least ordinarily) that I am caused to act by the reasons which favor doing so; it seems to be the case, rather, that *I* produce my decision *in view of* those reasons, and could have, in an unconditional sense, decided differently…Just as the non-Humean is apt to maintain that we not only perceive, e.g., the movement of the axe along with the separation of the wood, but the axe *splitting* the wood…, so I have the apparent perception of my actively and freely deciding to take Seneca Street to my destination and not Buffalo instead" (1995: 167-68). Richard Taylor, another leading agent-causal theorist, maintains that there are two introspective items of data: (1) That I *feel* that my behavior is sometimes the outcome of my deliberations, and (2) that in these and other cases, I *feel* that it is sometimes up to me what I do (1992, ch.5). He then concludes: "The only conception of action that accords with our data is one according to which people—and perhaps some other things too—are sometimes, but of course not always, self-determining beings; that is, beings that are sometimes the cause of their own behavior" (1992: 51).

[2] As J. Bargh and M. Ferguson point out, "Willfulness is assumed to reside in consciousness, and, therefore, a lack of conscious involvement in a process implies it was not [freely] willed" (2000: 925-26). See also Bargh 1989, 1996.

[3] To O'Connor's credit, he recognizes this point. He states: "It is highly plausible that this self-determining capacity strictly requires conscious awareness. This appears to follow from the very way in which active power has been characterized as structured by motivating reasons and as allowing the free formation of executive states of intention in accordance with one of the possible courses of action represented to oneself. (I am tempted to think that one should be able to explicitly demonstrate the absurdity of supposing an agent-causal capacity as being exercised entirely unconsciously)" (2000: 122).

[4] I believe this is part of the story, yet not the whole story, of why we feel free. I do not think it is the *whole* story because it does not explain why, upon introspection, we also feel the positive power of active freedom and self-determination. Not

being aware of deterministic causes may explain why we believe no such causes exist, but it does not fully explain the phenomenology of agent-causation. To fully explain why we feel free, I believe we will also have to examine the phenomenology surrounding our feeling of "self-causation" and our feeling of "intentional control" over our actions. The *apparent* transparency and infallibility of consciousness, I maintain, is one of four main phenomenological features of consciousness that create the illusion of free will. For a full account of how the cognitive illusion of free will is created, see my *Free Will and Conscious Illusion: A Determinist Account of the Cognitive Illusion of Free Will* (in progress).

[5] I have elsewhere called this the "Cartesian assumption" and have presented arguments against it. See Caruso, 2005.

[6] In fact, the majority of philosophers, *pace* John Searle (1990, 1992), now agree that there are nonconscious intentional states.

[7] As Timothy Wilson puts it, the view "that consciousness is the tip of the mental iceberg…[was] short of the mark by quite a bit—it may be more the size of a snowball on top of that iceberg" (2002: 6).

[8] See, Wilson, 2002; Bargh & Ferguson, 2000; Bargh & Chartrand, 1999; Bargh, 1990, 1994, 1996, 1997; Bargh, Barndollar, Gollwitzer, & Trotschel, 2001; Chartrand & Bargh, 1996; Bechara, H. Damsio, Tranel, & A. Damasio, 1997; Dijksterhuis, 2004; Dijksterhuis, Knippenberg, Spears, Postmes, Stapel, Koomen, & Scheepers, 1998; Aarts & Dijksterhuis, 2000, 2003; Dijksterhuis & Knippenberg, 1998; Steele & Aronson, 1995, and Libet, 1985, 1987, 1993.

[9] I should point out that when talking about nonconscious mental states, I do not mean simply to be talking about dispositional states: states that are disposed to be occurrent conscious states. I mean to be making the stronger claim that these are occurrent nonconscious states—states that influence behavior and interact with other mental states, both conscious and nonconscious.

[10] Timothy Wilson has even compared it to "Descartes' error" of Cartesian dualism. He writes: "Descartes made a related error that is less well known but no less egregious. Not only did he endow the mind with a special status that was unrelated to physical laws; he also restricted the mind to consciousness. The mind consists of all that people consciously think, he argued, and nothing else. This equation of thinking and consciousness eliminates, with one swift stroke, any possibility of nonconscious thought—a move that was called the 'Cartesian catastrophe' by Arthur Koestler and 'one of the fundamental blunders made by the human mind' by Lancelot Whyte. Koestler rightly notes that this idea led to 'an impoverishment of psychology which it took three centuries to remedy'" (2002: 9-10).

[11] Although Descartes famously entertained the possibility that the content of my mental states may not match reality, he never entertained the possibility that my own mental states may diverge from my conscious awareness of them. Descartes claims in numerous places that one cannot be mistaken about how things *seem* to them to be in consciousness (see, e.g., CSM II, 19; AT VII, 29). For Descartes, our judgments about our mental states are certain and infallible.

[12] As Rosenthal points out: "[C]onsciousness does not always represent our

mental states accurately. Consciousness seems infallible because it never shows itself to be mistaken and it's tempting to think that there's no other way to know what mental states one is in. But consciousness is not the only way to determine what mental state one is in, and there is sometimes compelling independent evidence that goes against what consciousness tell us" (Rosenthal, 2004: 27).

[13] For a good review, see Wilson 2003.

[14] See Wilson (2003) for more on the limits of introspective reports. With regard to emotions research, for example, Wilson writes: "With the advent of implicit measures and new theories about brain functioning, many researchers now argue that some feelings are inaccessible to conscious awareness, not necessarily because of repression, but because of the architecture of the mind" (Wilson, 2003: 134; see also, LeDoux, 1996; Wilson, 2002).

[15] For a good review of confabulation, which includes examples from split-brain patients, people suffering from organic amnesia, and people acting out posthypnotic suggestion, see Wilson, 2002: ch.5.

[16] In their original paper, Nisbett and Wilson (1977) argued that people often make inaccurate reports about the causes of their responses because there is little or no introspective access to higher order cognitive processes. They theorized that when people try to give introspective reports on the causes of their behavior, what they are really doing is making reasonable inferences about what the causes must have been, not giving direct introspective reports of the actual causes. A number of critics accurately pointed out that this thesis was far too extreme (Smith & Miller, 1978; Ericsson & Simon, 1980; Gavanski & Hoffman, 1987), and Wilson has since modified his views (see, Wilson & Stone, 1985; Wilson, 2002). My position is that we often do have direct access to our own mental states (i.e., our higher cognitive processes), but that the way we are aware of such states allows for the possibility of misrepresentation and confabulation. I further maintain that we need to distinguish between ordinary consciousness and introspective consciousness. I believe that error can occur at both levels and that people often conflate the two types of mistakes