Précis of *The illusion of conscious will*

Daniel M. Wegner
Department of Psychology, Harvard University, Cambridge, MA 02138.
wegner@wjh.harvard.edu http://www.wjh.harvard.edu/~wegner/

Abstract: The experience of conscious will is the feeling that we are doing things. This feeling occurs for many things we do, conveying to us again and again the sense that we consciously cause our actions. But the feeling may not be a true reading of what is happening in our minds, brains, and bodies as our actions are produced. The feeling of conscious will can be fooled. This happens in clinical disorders such as alien hand syndrome, dissociative identity disorder, and schizophrenic auditory hallucinations. And in people without disorders, phenomena such as hypnosis, automatic writing, Ouija board spelling, water dowsing, facilitated communication, speaking in tongues, spirit possession, and trance channelling also illustrate anomalies of will – cases when actions occur without will or will occurs without action. This book brings these cases together with research evidence from laboratories in psychology and neuroscience to explore a theory of apparent mental causation. According to this theory, when a thought appears in consciousness just prior to an action, is consistent with the action, and appears exclusive of salient alternative causes of the action, we experience conscious will and ascribe authorship to ourselves for the action. Experiences of conscious will thus arise from processes whereby the mind interprets itself – not from processes whereby mind creates action. Conscious will, in this view, is an indication that we think we have caused an action, not a revelation of the causal sequence by which the action was produced.

Keywords: apparent mental causation; automatism; conscious will; determinism; free will; perceived control

1. The illusion (Ch. 1)

So, here you are, reading about conscious will. How could this have happened? One way to explain it would be to examine the causes of your behavior. A team of scientists could study your reported thoughts, emotions, and motives, your genetics and your history of learning, experience, and development, your social situation and culture, your memories and reaction times, your physiology and neuroanatomy, and lots of other things as well. If they somehow had access to all the information they could ever want, the assumption of psychology is that they could uncover the mechanisms that give rise to all your behavior, and so could certainly explain why you are reading these words at this moment. However, another way to explain the fact of your reading these lines is just to say that you decided to begin reading. You consciously willed what you are doing.

The ideas of conscious will and psychological mechanism have an oil and water relationship, having never been properly reconciled. One way to put them together is to say that the mechanistic approach is the explanation preferred for scientific purposes, but that the person’s experience of conscious will is utterly convincing and important to the person – and so must be understood scientifically as well. The mechanisms underlying the experience of will are themselves a fundamental topic of scientific study.

1.1. Conscious will

Conscious will is usually understood in one of two ways. It is common to talk about conscious will as something that is experienced when we perform an action: Actions feel willed or not, and this feeling of voluntariness or doing a thing “on purpose” is an indication of conscious will. It is also common, however, to speak of conscious will as a force of mind, a name for the causal link between our minds and our actions. One might assume that the experience of consciously willing an action and the causation of the action by the person’s conscious mind are the same thing. As it turns out, however, they are entirely distinct, and the tendency to confuse them is the source of the illusion of conscious will. So, to begin, we will need to look into each in turn, first examining will as an experience and then considering will as a causal force.

Wegner: Précis of The illusion of conscious will

nal. This definition puts the person's experience at the very center of the whole concept—the will is not some cause or force or motor in a person, but rather is the personal conscious feeling of such causing, forcing, or motoring. Hume's definition makes sense because the occurrence of this conscious experience is an absolute must for anyone to claim to have done something that he or she consciously willed.

Without an experience of willing, even actions that look entirely voluntary from the outside still fall short of qualifying as truly willed. Intentions, plans, and other thoughts can be experienced, and still the action is not willed if the person says it was not. If a person plans to take a shower, for example, and says that she intends to do it as she climbs into the water, spends 15 minutes in there scrubbing up nicely, and then comes out reporting that she indeed seems to have had a shower—but yet also reports not feeling she had consciously willed her showering—who are we to protest? Consciously willing an action requires a feeling of doing (Ansfield & Wegner 1996), a kind of internal "oomph" that somehow certifies authentically that one has done the action. If the person did not get that feeling about her shower, then even if we climbed in with her to investigate, there is no way we could establish for sure whether she consciously willed her showering.

The fact that experiences of conscious will can only be established by self-reports ("I showered, yes I did") would be quite all right if the self-reports always corresponded with some other outward indication of the experience. However, this correspondence does not always happen. The experience of will that is so essential for the occurrence of consciously willed action does not always accompany actions that appear by other indications to be willed.

Consider, for example, the case of people who have alien hand syndrome, a neuropsychological disorder in which a person experiences one hand as operating with a mind of its own. Alien hand patients typically experience one hand as acting autonomously. They do not experience willing its actions, and may find it moving at cross-purposes with their conscious intention. This syndrome is often linked with damage to the middle of the frontal lobes on the side of the brain opposite the affected hand (Gasquoine 1993). Banks and colleagues (1989) report an alien hand patient whose left hand would tenaciously grope for and grasp any nearby object, pick and pull at her clothes, and even grab her throat during sleep.... She slept with the arm tied to prevent nocturnal misbehavior. She never denied that her left arm and hand belonged to her, although she did refer to her limb as though it were an autonomous entity. (Banks et al. 1989, p. 456)

Should the alien hand's movements be classed as willed or unwilled? On the one hand (pun can't be helped), the alien hand seems to do some fairly complicated things, acts we might class as willful and voluntary if we were just watching and hadn't learned of the patient's lamentable loss of control. In the case of another patient, for example,

While playing checkers on one occasion, the left hand made a move he did not wish to make, and he corrected the move with the right hand; however, the left hand, to the patient's frustration, repeated the false move. On other occasions, he turned the pages of the book with one hand while the other tried to close it; he shaved with the arm tied to prevent nocturnal misbehavior. She never denied that her left arm and hand belonged to her, although she did refer to her limb as though it were an autonomous entity. (Banks et al. 1989, p. 456)

By the looks of it, the alien hand is quite willful. On the other hand (as the pun drags on), however, the patient does not experience these actions as consciously willed.

Brain damage is not the only way that the experience of will can be undermined. Consider, for instance, the feelings of involuntariness that occur during hypnosis. Perhaps the most profound single effect of hypnosis is the feeling that your actions are happening to you, rather than that you are doing them (Lynn et al. 1990). To produce this experience, a hypnotist might suggest, "Please hold your arm out to your side. Now, concentrate on the feelings in your arm. What you will find is that your arm is becoming heavy. It feels as though a great weight were pulling it down. It is so very heavy. It is being pulled down, down toward the ground. Your arm is heavy, very heavy. It is getting so heavy you can't resist. Your arm is falling, falling downward toward the ground." With enough of this patter, many listeners will indeed experience the arm becoming heavy, and some will even find their arm falling down. When quizzed on it, these individuals often report that they felt no sense of moving their arm voluntarily, but rather experienced the downward movement as something that happened to them. This does not occur for everyone in this situation, only some proportion, but it nonetheless indicates that the experience of will can be manipulated in a voluntary action.

In the case of hypnotic involuntariness, the person has a very clear and well-rehearsed idea of the upcoming action. Admittedly, this idea of the action is really phrased more as an expectation ("My arm will fall") than as an intention ("I will lower my arm"), but it nonetheless occurs before the action when an intention normally happens, and it provides a distinct preview of the action that is to come (Kirsch & Lynn 1998b, Spanos 1986b). Hypnotic involuntariness thus provides an example of the lack of experience of will that is yet more perplexing than alien hand syndrome. With alien hand, the person simply does not know what the hand will do, but with hypnosis, conscious will is lacking— even when knowledge of the action is present. And without the experience of willing, a person is not likely to move the action into the "consciously willed" category. If it does not feel as though you did it, then it does not seem that the will was operating.

Another case of the absence of an experience of will occurs in table-turning, a curious phenomenon discovered in the spiritualist movement in Europe and the United States in the mid-nineteenth century (Ansfield & Wegner 1996; Carpenter 1888). To create this effect, a group of people sits around a table with their hands on its surface. If they are convinced that the table might move as the result of spirit intervention (or if they are even just hoping for such an effect), and sit patiently waiting for such movement, it is often found that the table does start to move after some time. It might even move about the room or begin rotating so quickly that the participants can barely keep up. Carpenter (1888, pp. 292–93) observed "all this is done, not merely with the least consciousness on the part of the performers that they are exercising any force of their own, but for the most part under the full conviction that they are not." Incidentally, table-turning was sufficiently controversial that it attracted the attention of the chemist and physicist Michael Faraday, who proceeded to test the source of the table movement. He placed force measurement devices between participants' hands and the table, and found that the
source of the movement was their hands and not the table (Faraday 1853).

Such examples of the separation of action from the experience of will suggest that it is useful to draw a distinction between them. Table 1 shows four basic conditions of human action – the combinations that arise when we emphasize the distinction between action and the sense of acting willfully. The upper left corner contains the expected correspondence of action and the feeling of doing – the case when we do something and feel also that we are doing it. This is the noncontroversial case, or perhaps the assumed human condition. The lower right corner is also noncontroversial, the instance when we are not doing anything and feel we are not.

The upper right – the case of no feeling of will when there is in fact the occurrence of action – encompasses the examples we have been inspecting thus far. The movement of alien hands, the case of hypnotic suggestion of arm heaviness, and table-turning all fit this quadrant, as they involve no feeling of doing in what appear otherwise to be voluntary actions. These can be classed in general as automatisms. The other special quadrant of the table includes cases of the illusion of control. Ellen Langer (1975) used the term to describe instances when people have the feeling that they are doing something when they actually are not doing anything.

The illusion of control is acute in our interactions with machines – as when we do not know whether our push of an elevator button or Coke machine selection has done anything, yet sense that it has. The illusion is usually studied with judgments of contingency (e.g., Matute 1996) by having people try to tell whether they are causing a particular effect, for example, turning on a light, by doing something, such as pushing a button, when the button and the light are not perfectly connected and the light may flash randomly by itself. But we experience the illusion, too, when we roll dice or flip coins in a certain way, hoping that we will thus be able to influence the outcome. It even happens sometimes that we feel we have contributed to the outcome of a sporting event on TV just by our presence in the room (“Did I just jinx them by running off to the fridge?”).

Most of the things we do in everyday life seem to fall along the “normal” diagonal in this fourfold table. Action and the experience of will usually correspond, so we feel we are doing things willfully when we actually do them, and feel we are not doing something when in truth we have not done it. Still, the automatisms and illusions of control that lie off this diagonal remind us that action and the feeling of doing are not locked together inevitably. They come apart often enough that one wonders whether they may be produced by separate systems in the mind. The processes of mind that produce the experience of will may be quite distinct from the processes of mind that produce the action itself. As soon as we accept the idea that the will should be understood as an experience of the person who acts, we come to realize that conscious will is not inherent in action – there are actions that have it and actions that do not.

1.1.2. The force of conscious will. Will is not only an experience, but also a force. Because of this, it is tempting to think that the conscious experience of will is a direct perception of the force of will. The feeling that one is purposefully not having a cookie, for example, can easily be taken as an immediate perception of one’s conscious mind causing this act of self-control. We seem to experience the force within us that keeps the cookie out of our mouths, but the force is not the same thing as the experience.

When conscious will is described as a force, it can take different forms. Will can come in little dabs to produce individual acts, or it can be a more long-lasting property of a person, a kind of inner strength or resolve. Just as a dish might have hotness or an automobile might have the property of being red, a person seems to have will, a quality of power that causes his or her actions. The force may be with us. Such will can be strong or weak, and so can serve to explain things such as one person’s steely persistence in the attempt to dig a swimming pool in the back yard, for example, or another person’s knee-buckling weakness for chocolate. The notion of strength of will has been an important intuitive explanation of human behavior since the ancients (Charlton 1988), and it has served throughout the history of psychology as the centerpiece of the psychology of will.

The classic partition of the mind into three functions includes cognition, emotion, and conation – the will or volitional component (e.g., James 1890).

The will in this traditional way of thinking is an explanatory entity of the first order. In other words, it explains lots of things but nothing explains it. As Joseph Buchanan (1812) described it, “Volition has commonly been considered by metaphysical writers, as consisting in the exertion of an innate power, or constituent faculty of the mind, denominated will, concerning whose intrinsic nature it is fruitless and unnecessary to inquire” (p. 298). At the extreme, of course, this view of the will makes the scientific study of it entirely out of the question, and suggests instead that it ought to be worshiped. Pointing to will as a force in a person that causes the person’s action is the same kind of explanation as saying that God has caused an event. This is a stopper that trumps any other explanation, but that still seems not to explain anything at all in a predictive sense. Just as we cannot tell what God is going to do, we cannot predict what the will is likely to do either.

The notion that will is a force residing in a person has a further problem. Hume remarked on this when he described the basic difficulty that occurs whenever a person perceives causality in an object. Essentially, he pointed out that causality is not a property inhering in objects. For instance, when we see a bowling ball go scooting down the lane and smashing into the pins, it certainly seems as though the ball has some kind of causal force in it. The ball is the cause and the explosive reaction of the pins is the effect. Hume pointed out, though, that you cannot see causation in something, but must only infer it from the constant relation between cause and effect. Every time the ball rolls into the pins, they bounce away. Ergo, the ball caused the pins to move. But there is no property of causality nestled somewhere in that ball, or hanging somewhere in space between the ball and pins, that somehow works this magic. Causation is an event, not a thing or a characteristic or attribute of an object.
Hume realized, then, that calling the will a force in a person’s consciousness— even in one’s own consciousness—must always overreach what we can see (or even introspect), and so should be understood as an attribution or inference. This is not to say that the concept of will power is useless. Rather, Hume’s analysis suggests that the concepts of force of will or will power must be accompanied by careful causal inference. These ideas can be used as the basis for scientific theories of human behavior, certainly, as they serve as summaries of the degree of relationship that may exist between the mind and behavior. But we must be careful to distinguish between such empirical will—the causality of the person’s conscious thoughts as established by a scientific analysis of their covariation with the person’s behavior—and the phenomenal will—the person’s reported experience of will. The empirical will can be measured by examining the degree of covariation between the person’s self-reported conscious thought and the person’s action, and by assessing the causal role of that thought in the context of other possible causes of the action (and possible causes of the thought as well).

The empirical will—the actual relationship between mind and action—is a central topic of scientific psychology. In psychology, clear indications of the empirical will can be found whenever causal relationships are observed between people’s thoughts, beliefs, intentions, plans, or other conscious psychological states and their subsequent actions. The feeling of consciously willing our action, in contrast, is not a direct readout of such scientifically verifiable will power. Rather, it is the result of a mental system whereby each of us estimates moment-to-moment the role that our minds play in our actions.

1.2. Mind perception

Why would people mistake the experience of will for a causal mechanism? Why is it that the phenomenal will so easily overrides any amount of preaching by scientists about the mechanisms underlying human action? Now as a rule, when people find an intuition so wildly intriguing that they regularly stand by it and forsake lots of information that is technically more correct, they do so because the intuition fits. It is somehow part of a bigger scheme of things that they simply cannot discard. So, for example, people once held tight to the Ptolemaic idea that the sun revolves around the earth, in part because this notion fit their larger religious conception of the central place of the earth in God’s universe. In exactly this way, conscious will fits a larger conception—our understanding of causal agents.

1.2.1. Causal agency. Most adult humans have a very well-developed idea of a particular sort of entity, an entity that does things. We appreciate that a dog, for example, will often do things that are guided not by standard causal principles, but rather by a teleological or purposive purpose. Dogs often seem to be goal-oriented, as they behave in ways that only seem to be understandable in terms of goals (including some fairly goofy ones, yes, but goals nonetheless). They move toward things that they subsequently seem to have wanted (because they consume them or sniff them), and they move away from things that we can imagine they might not like (because the things are scary or loud or seem to be waving a rolled-up newspaper). Dogs, like horses and fish and crickets and even some plants, seem to be understandable through a special kind of thinking about goal-oriented entities that does not help us at all in thinking about bricks, buttons, or other inanimate objects.

The property of goal seeking is not just something we attribute to living things, as we may appreciate this feature in computers or robots or even thermostats. But the important characteristic of such goal-seeking entities is that we understand them in terms of where we think they are headed rather than in terms of where we think they have been. Unlike a mere object, which moves or “acts” only when it has been caused to do so by some prior event, a causal agent moves or acts apparently on its own, in the pursuit of some future state—the achievement of a goal. Fritz Heider (1958; Heider & Simmel 1944) observed that people perceive persons as causal agents—origins of events—and that this is the primary way in which persons are understood in a manner that physical objects and events are not.

Causal agency, in sum, is an important way in which people understand action, particularly human action. In the process of understanding actions performed by oneself or by another, the person will appreciate information about intentions, beliefs, desires, and plans, and will use this information in discerning just what the agent is doing. The intuitive appeal of the idea of conscious will can be traced in part to the embedding of the experience of will, and of the notion that will has a force, in the larger conception of causal agency. Humans appear to be goal-seeking agents who have the special ability to envision their goals consciously in advance of action. The experience of conscious will feels like being a causal agent.

1.2.2. Mechanisms and minds. We all know a lot about agents and goals and desires and intentions, and use these concepts all the time. These concepts are only useful, however, for understanding a limited range of our experience. The movements of clock hands and raindrops and electric trains, for example, can be understood in terms of causal relations that have no consciousness or will at all. They are mechanisms. Extending the notion of causal agency to these items—to say these things have the ability to cause themselves to be—he does not fit very well with the physical causal relations we perceive all around us. Imagine for a moment a spoon, knife, and fork deciding to go for a walk.
to the far end of the dinner table ("we’re off to see the salad . . ."), and you can see the problem. Things do not usually will themselves to move, whereas people seem to do this all the time.

This rudimentary observation suggests that people have at hand two radically different systems of explanation, one for minds and one for everything else. Mentalistic explanation works wonders for understanding minds, but it does not work elsewhere — unless we want to start thinking that everything from people to rocks to beer cans to the whole universe actually does what it consciously wants. Mechanistic explanation, in turn, is just splendid for understanding those rocks and beer cans, not to mention the movements of the planets, but meanwhile leaves much to be wanted in understanding minds.

Each of us is quite comfortable with using these two very different ways of thinking about and explaining events — a physical, mechanical way and a psychological, mental way. In the mechanical explanatory system, people apply intuitive versions of physics to questions of causality, and so they think about cause and effects as events in the world. In the mental explanatory system, in turn, people apply implicit psychological theories to questions of causality, focusing on issues of conscious thoughts and the experience of will as they try to explain actions. In the mechanical way of thinking, all the psychological trappings are unnecessary; a physical system such as a clock, for example, does not have to intend to keep time or to experience doing so. The essence of the mental explanatory system, in contrast, is the occurrence of the relevant thoughts and feelings about the action. In this system, the objects and events of physical causality are not particularly important; a person might experience having willed the death of an enemy and become wracked with guilt, for example, even though there was no mechanism for this to have happened.

These two explanatory systems fall into place as children develop ways of understanding both the physical and psychological worlds. The first inklings that mind perception and mechanistic explanation might develop separately in children came from Jean Piaget, whose perspective has culminated in the contemporary literature on the development of "theory of mind" in animals (Premack & Woodruff 1978) and in children (e.g., Wellman 1992), and in work that contrasts how children develop an understanding of agency, intention, and will with how they develop an understanding of causality, motion, and the principles of physics (e.g., Carey 1996; Gelman et al. 1995). Neither the perception of the physical world nor the perception of the mental world is a "given" to the new human. Although the neonate has rudimentary abilities in both areas, both systems must be developed over time and experience as ways of understanding what all is going on.

The idea that mind perception is variable has also been noted by Dennett (1987; 1996), who captured this observation in suggesting that people take an "intentional stance" in perceiving minds that they do not take in perceiving most of the physical world. The degree to which we perceive mindedness in phenomena can change, such that under some circumstances we see our pet pooch as fully conscious and masterfully deciding just where it would be good to scratch himself, whereas under other circumstances we may have difficulty extending the luxury of presumed conscious thought and human agency even to ourselves. It is probably the case, too, that the degree of mechanical causality we perceive is something that varies over time and circumstance. Viewing any particular event as mentally or mechanically caused, therefore, can depend on a host of factors and can influence dramatically how we go about making sense of it. And making sense of our own minds as mentally causal systems — conscious agents — includes accepting our feelings of conscious will as authentic.

1.3. Real and apparent mental causation

Any magician will tell you the key to creating a successful illusion: The illusionist must make a marvelous, apparently magical event into the easiest and most immediate way to explain what are really mundane events. Kelley (1980) described this in his analysis of the underpinnings of magic in the perception of causality. He observed that stage magic involves a perceived causal sequence — the set of events that appears to have happened — and a real causal sequence — the set of events the magician has orchestrated behind the scenes. The perceived sequence is what makes the trick. Laws of nature are broken willy-nilly as people are sawed in half, birds and handkerchiefs and rabbits and canes and what-have-you appear from nothing, and also disappear, or for that matter turn into each other and then back again.

The real sequence is often more complicated or unexpected than the illusion, but many of the real events are not perceived. The magician needs special pockets, props, and equipment, and develops wiles to misdirect audience attention from the real sequence. In the end, the audience observes something that seems to be simple, but in fact it may have been achieved with substantial effort, preparation, practice, and thought on the magician's part. The lovely assistant in a gossamer gown apparently flying effortlessly on her back during the levitation illusion is in fact being held up by a 600-pound pneumatic lift hidden behind the specially rigged curtain. It is the very simplicity of the illusory sequence, the shorthand summary that circumvents all the poor magician's toil, which makes the trick so compelling. The lady levitates. The illusion of conscious will occurs by much the same technique (Wegner 2003a).

The real causal sequence underlying human behavior involves a massively complicated set of mechanisms. Everything that psychology studies can come into play to predict and explain even the most innocuous wink of an eye, not to mention some of the more lengthy and elaborate behaviors of which humans are capable. Each of our actions is really the culmination of an intricate set of physical and mental processes, including psychological mechanisms that correspond to the traditional concept of will — in that they involve linkages between our thoughts and our actions. This is the empirical will. However, we do not see this. Instead, we readily accept the far easier explanation of our behavior that our Houdini-esque minds present to us: We think we did it.

Science fiction writer Arthur C. Clarke (1973, p. 21) remarked that "Any sufficiently advanced technology is indistinguishable from magic." Clarke meant this to refer to the fantastic inventions we might discover in the future, or might find if we were to travel to advanced civilizations. However, the insight also applies to self-perception. When we turn our attention to our own minds, we find that we are suddenly faced with trying to understand an unimaginably advanced technology. We cannot possibly know (let alone keep track of) the tremendous number of mechanistic influences on our behavior, because we have the fortune of...
inhabiting some extraordinarily complicated machines. So we develop a shorthand—a belief in the causal efficacy of our conscious thoughts. We believe in the magic of our own causal agency.

The mind creates this continuous illusion because it really doesn’t know what causes its actions. Whatever empirical will there is is rumbling along in the engine room—an actual relation between thought and action—might in fact be totally inscrutable to the conscious mind. The mind has a self-explanation mechanism that produces a roughly continuous sense that what is in consciousness is the cause of action—the phenomenal will—whereas in fact the mind actually cannot ever know itself well enough to be able to say what the causes of its actions are. To quote Spinoza in The Ethics: “Men are mistaken in thinking themselves free; their opinion is made up of consciousness of their own actions, and ignorance of the causes by which they are determined. Their idea of freedom, therefore, is simply their ignorance of any cause for their actions” (Spinoza 1677/1883, Part II, p. 105). In the more contemporary phrasing of Minsky (1985, p. 306), “none of us enjoys the thought that what we do depends on processes we do not know; we prefer to attribute our choices to colition, will, or self-control... Perhaps it would be more honest to say, ‘My decision was determined by internal forces I do not understand’” (italics in original).

2. Apparent mental causation (Ch. 3)

Imagine for a moment that by some magical process, you could always know when a particular tree branch would move in the wind. Just before it moved, you knew it was going to move, in which direction, and just how it would do it. Not only would you know this, but let us assume that the same magic would guarantee that you would happen to be thinking about the branch just before each move. You would look over, and then just as you realized it was going to move, it would do it! In this imaginary situation, you could eventually come to think that you were somehow causing the movement. You would seem to be the source of the distant branch’s action, the agent that wills it to move. The feeling that one is moving the tree branch surfaces in the same way that one would get the sense of performing any action at a distance. All it seems to take is the appropriate foreknowledge of the action. Indeed, with proper foreknowledge it is difficult not to conclude one has done the act, and the feeling of doing may well-up in direct proportion to the perception that relevant ideas had entered one’s mind before the action. This is beginning to sound like a theory.

2.1. A theory of apparent mental causation

The experience of will may be a result of the same mental processes that people use in the perception of causality more generally. The theory of apparent mental causation, then, is this: people experience conscious will when they interpret their own thought as the cause of their action (Wegner & Wheatley 1999). This means that people experience conscious will quite independently of any actual causal connection between their thoughts and actions. Reductions in the impression that there is a link between thought and action may explain why people get a sense of involuntariness even for actions that are voluntary, for example, during motor automatisms such as table-turning, or in hypnosis, or in psychologically disordered states such as dissociation. And inflated perceptions of the link between thought and action, in turn, may explain why people experience an illusion of conscious will at all.

The person experiencing will, in this view, is in the same position as someone perceiving causation as one billiard ball strikes another. As we learned from Hume, causation in bowling, billiards, and other games is inferred from the constant conjunction of ball movements. It makes sense, then, that will—an experience of one’s own causal influence—is inferred from the conjunction of events that lead to action. Now, in the case of billiard balls, the players in the causal analysis are quite simple: one ball and the other ball. One rolls into the other and a causal event occurs. What are the items that seem to click together in our minds to yield the perception of will?

One view of this was provided by Ziehen (1999), who suggested that thinking of self before action yields the sense of agency. He proposed that we finally come to regard the ego-idea as the cause of our actions because of its very frequent appearance in the series of ideas preceding each action. It is almost always represented several times among the ideas preceding the final movement. But the idea of the relation of causality is an empirical element that always appears when two successive ideas are very closely associated (Ziehen 1999, p. 296).

And indeed, there is evidence that self-attention is associated with perceived causation of action. People in an experiment by Duval and Wicklund (1973) were asked to make attributions for hypothetical events (a hypothetical item: “Imagine you are rushing down a narrow hotel hallway and bump into a housekeeper who is backing out of a room”). When asked to decide who was responsible for such events, they assigned more causality to themselves if they were making the judgments while they were self-conscious. Self-consciousness was manipulated in this study by having the participants sit facing a mirror, but other contrivances—such as showing people their own video image or having them hear their tape-recorded voice—also enhanced causal attribution to self (Gibbons 1990).

This tendency to perceive oneself as causal when thinking about oneself is a global version of the more specific process that appears to underlie apparent mental causation. The specific process is the perception of a causal link not only between self and action, but between one’s own thought and action. We tend to see ourselves as the authors of an act when we have experienced relevant thoughts about the act at an appropriate interval in advance, and so can infer that our own mental processes have set the act in motion. Actions we perform that are not presaged in our minds, in turn, would appear not to be caused by our minds. The intentions we have to act may or may not be causes, but this does not matter, as it is only critical that we perceive them as causes if we are to experience conscious will.

In this analysis, the experience of will is not a direct read-out of some psychological force that causes action from inside the head. Rather, will is experienced as a result of an interpretation of the apparent link between the conscious thoughts that appear in association with action and the nature of the observed action. Will is experienced as the result of self-perceived apparent mental causation. Thus, in line with facets of several existing theories (Brown 1989; Clax-
2.2. Principles of causal inference

How do we go about drawing the inference that our thought has caused our action? Several ideas about this pop up on considering the tree branch example once more. Think, for instance, of what could spoil the feeling that you had moved the branch. If the magic limb moved before you thought of it moving, there would be nothing out of the ordinary and you would experience no sense of willful action. The thought of movement would be interpretable as a memory or even a perception of what had happened. If you thought of the tree limb moving and then something quite different moved (say, a nearby chicken dropped to its knees), again there would be no experience of will. The thought would be irrelevant to what had happened, and you would see no causal connection. And if you thought of the tree limb moving but noticed that something other than your thoughts had moved it (say, a squirrel), no will would be sensed. There would simply be the perception of an external causal event. These observations point to three key sources of the experience of conscious will: the priority, consistency, and exclusivity of the thought about the action (Wegner & Wheatley 1999). For the perception of apparent mental causation, the thought should occur before the action, be consistent with the action, and not be accompanied by other potential causes.

Studies of how people perceive external physical events (Michotte 1963) indicate that the perception of causality is highly dependent on these features of the relationship between the potential cause and potential effect. The candidate for the role of cause must come first or at least at the same time as the effect, it must yield movement that is consistent with its own movement, and it must be unaccompanied by rival causal events. The absence of any of these conditions tends to undermine the perception that causation has occurred. Similar principles have been derived for the perception of causality for social and everyday events (Einhorn & Hogarth 1986; Gilbert 1997; Kelley 1972; McClure 1998), and have also emerged from analyses of how people and other organisms respond to patterns of stimulus contingency when they learn (Alloy & Tabachnik 1984; Young 1985). The application of these principles to the experience of conscious will can explain phenomena of volition across a number of areas of psychology.

2.3. Intentions as previews

The experience of will is the way our minds portray their operations to us, not their actual operation. Because we have thoughts of what we will do, we can develop causal theories relating those thoughts to our actions on the basis of priority, consistency, and exclusivity. We come to think of these prior thoughts as intentions, and we develop the sense that the intentions have causal force even though they are actually just previews of what we may do. Yet, in an important sense, it must be the case that something in our minds plays a causal role in making our actions occur. That something is, in the theory of apparent mental causation, a set of unconscious mental processes that cause the action. At the same time, that “something” is very much like the thoughts we have prior to the action.

One possibility here is that thought and action arise from coupled unconscious mental systems. Brown (1989) has suggested that consciousness of an action and the performance of the action are manifestations of the same “deep structure.” In the same sense that the thought of being angry might reflect the same underlying process as the experience of facial flushing, the thought and performance of a
Wegner: Précis of The illusion of conscious will

voluntary action might be different expressions of a singular underlying system. The coupling of thought and action over time in the adult human is really quite remarkable if the thought is not causing the action, so there must be some way in which the two are in fact often connected.

The co-occurrence of thought and action may happen because thoughts are normally thrust into mind as *previews* of what will be done. The ability to know what one will do, and particularly to communicate this to others verbally, would seem to be an important human asset, something that promotes far more effective social interaction than might be the case if we all had no idea of what to expect of ourselves or of anyone around us. The thoughts we find coming to our minds in frequent coordination with what we do may thus be produced by a special system whose job it is to provide us with ongoing verbalizable previews of action. This preview function could be fundamentally important for the facilitation of social interaction. Intentions, in this analysis, are to action what turn signals are to the movements of motor vehicles. They do not cause the movements, they preview them.

By this logic, real causal mechanisms underlying behavior are never present in consciousness. Rather, the engines of causation operate without revealing themselves to us, and so may be unconscious mechanisms of mind. The research suggesting a fundamental role for automatic processes in everyday behavior (Bargh 1997) can be understood in this light. The real causes of human action are unconscious, so it is not surprising that behavior could often arise – as in automaticity experiments – without the person having conscious insight into its causation. Conscious will itself arises from a set of processes that are not the same processes as those that cause the behavior to which the experience of will pertains, however. So, even processes that are not automatic – mental processes described as “controlled” (Posner & Snyder 1975) or “conscious” (Wegner & Bargh 1998) – have no direct expression in a person’s experience of will. Such “controlled” processes may be less efficient than automatic processes and require more cognitive resources, but even if they occur along with an experience of control or conscious will, this experience is not a direct indication of their real causal influence. The experience of conscious will is just more likely to accompany inefficient processes than efficient ones because there is more time available prior to action for inefficient thoughts to become conscious, thus to prompt the formation of causal inferences linking thought and action. This might explain why controlled/conscious processes are often linked with feelings of will, whereas automatic processes are not. Controlled and conscious processes are simply those that lumber along so inefficiently that there is plenty of time for previews of their associated actions to come to mind and allow us to infer the operation of conscious will (Wegner, in press).

The unique human convenience of conscious thoughts that preview our actions gives us the privilege of feeling we willfully cause what we do. In fact, however, unconscious and inscrutable mechanisms create both conscious thought about action and the action as well, and also produce the sense of will we experience by perceiving the thought as cause of action. So, although our thoughts may have deep, important, and unconscious causal connections to our actions, the experience of conscious will arises from a process that interprets these connections, not from the connections themselves.

3. The mind’s compass (Ch. 9)

Does the compass steer the ship? In some sense, you could say that it does, because the pilot makes reference to the compass in determining whether adjustments should be made to the ship’s course. If it looks as though the ship is headed west into the rocky shore, a calamity can be avoided with a turn north into the harbor. But of course, the compass does not steer the ship in any physical sense. The needle is just gliding around in the compass housing, doing no actual steering at all. It is thus tempting to relegate the little magnetic pointer to the class of epiphenomena – things that do not really matter in determining where the ship will go.

Conscious will is the mind’s compass. As we have seen, the experience of consciously willing action occurs as the result of an interpretive system, a course-sensing mechanism that examines the relations between our thoughts and actions and responds with “I willed this” when the two correspond appropriately. This experience thus serves as a kind of compass, alerting the conscious mind when actions occur that are likely to be the result of one’s own agency. The experience of will is therefore an indicator, one of those gauges on the control panel to which we refer as we steer. Like a compass reading, the feeling of doing tells us something about the operation of the ship beneath us. But also like a compass reading, this information must be understood as a conscious experience, a candidate for the dreaded “epiphenomenon” label. Just as compass readings do not steer the boat, conscious experiences of will do not cause human actions.

Why is it that the conscious experience of will exists at all? Why, if this experience is not a sensation of the personal causation of action, would we even go to the trouble of having it? What good is an epiphenomenon? The answer to this question becomes apparent when we appreciate conscious will as a feeling that organizes and informs our understanding of our own agency. Conscious will is a signal with many of the qualities of an emotion, one that reverberates through the mind and body to indicate when we sense having authored an action. The idea that conscious will is an emotion of authorship moves beyond the standard way in which people have been thinking about free will and determinism and presses toward a useful new perspective.

3.1. Free will and determinism

A book called The Illusion of Conscious Will certainly gives the impression of being a poke in the eye for readers who believe in free will. It is perfectly reasonable to look at the title and think the book is all about determinism and that it will give the idea of free will no fair hearing at all. And, of course, the line of thought here does take a decidedly deterministic approach. For all this, though, our discussion has actually been about the experience of free will, examining at length when people feel it and when they do not. The special idea we have been exploring is to explain the experience of free will in terms of deterministic or mechanistic processes.

On the surface, this idea seems not to offer much in the way of a solution for the classic question of free will and determinism. How does explaining the feeling of will in terms of deterministic principles help us to decide which one is true? Most philosophers and people on the street see this
as a fight between two big ideas, and they call for a decision on which one is the winner. As it turns out, however, a decision is not really called for at all. The usual choice we are offered between these extremes is not really a choice, but rather a false dichotomy. It is like asking: Shall we dance, or shall we move about the room in time with the music? The dichotomy melts when we explain one pole of the dimension in terms of the other. Still, this does not sit well with anyone who is married to the standard version of the problem, so we need to examine just how this usual choice leads us astray.

3.1.1. The usual choice. Most of us think we understand the basic issue of free will and determinism. The question seems to be whether all our actions are determined by mechanisms beyond our control, or whether at least some of them are determined by our free choice. Described this way, many people are happy to side with one possibility or the other. There are those of us who side with free will, and thus view members of the opposition as nothing but robo-geeks, creatures who are somehow disposed to cast away the very essence of their humanity and embrace a personal identity as automatons. There are others of us, however, who opt for the deterministic stance, and thus view the opposition as little more than bad scientists, a cabal of confused mystics with no ability to understand how humanity fits into the grand scheme of things in the universe. Viewed in each others' eyes, everyone comes out a loser.

The argument between these two points of view usually takes a simple form: The robogeeks point to the array of evidence that human behavior follows mechanistic principles, taking great pride in whatever data or experiences accumulate to indicate that humans are predictable by the rules of science. Meanwhile, the bad scientists ignore all of this and simply explain that their own personal experience carries the day. They know they have conscious will. And no one wins the argument. The usual clash falls on both sides because free will is a feeling, whereas determinism is a process. They are incommensurable. Free will is apples and determinism is oranges.

The illogic of treating free will and determinism as equal opposites becomes particularly trenchant when we try to make free will do determinism's causal job. What if, for example, we assume that free will is just like determinism, in that it is also a process whereby human behavior can be explained? Rather than all the various mechanistic engines that psychologists have invented or surmised in humans that might cause their behavior, imagine instead a person in whose head there is installed a small unit called the Free Willer. This is not the usual psychological motor, the bundle of thoughts or motives or emotions or neurons or genes – instead, it is a black box that just does things. Many kinds of human abilities and tendencies can be modeled in artificially intelligent systems, after all, and it seems on principle that we should be able to design at least the rudiments of a social intelligence as automatons. Then, there are others of us, however, who opt for the deterministic stance, and thus view the opposition as little more than bad scientists, a cabal of confused mystics with no ability to understand how humanity fits into the grand scheme of things in the universe. Viewed in each others' eyes, everyone comes out a loser.

The argument between these two points of view usually takes a simple form: The robogeeks point to the array of evidence that human behavior follows mechanistic principles, taking great pride in whatever data or experiences accumulate to indicate that humans are predictable by the rules of science. Meanwhile, the bad scientists ignore all of this and simply explain that their own personal experience carries the day. They know they have conscious will. And no one wins the argument. The usual clash falls on both sides because free will is a feeling, whereas determinism is a process. They are incommensurable. Free will is apples and determinism is oranges.

The illogic of treating free will and determinism as equal opposites becomes particularly trenchant when we try to make free will do determinism's causal job. What if, for example, we assume that free will is just like determinism, in that it is also a process whereby human behavior can be explained? Rather than all the various mechanistic engines that psychologists have invented or surmised in humans that might cause their behavior, imagine instead a person in whose head there is installed a small unit called the Free Willer. This is not the usual psychological motor, the bundle of thoughts or motives or emotions or neurons or genes – instead, it is a black box that just does things. Many kinds of human abilities and tendencies can be modeled in artificially intelligent systems, after all, and it seems on principle that we should be able to design at least the rudiments of a psychological process that has the property of freely willing actions.

But what exactly do we install? If we put in a module that creates actions out of any sort of past experiences or memories, that fashions choices from habits or attitudes or inherited tendencies, we do not get freedom, we get determinism. The Free Willer must be a mechanism that is unresponsive to any past influence. In Elbow Room: The Varieties of Free Will Worth Wanting, Dennett (1984) illustrated how hollow and unsatisfying free will of this kind might be. In essence, any such system makes sense only if it inserts some fickle indeterminacy into the person's actions. Dennett points out that it is not particularly interesting or fun to have a coin flipper added to the works somewhere between “sensory input” and “behavior output.” Who would want free will if it is nothing more than an internal coin flip? This is not what we mean when we talk about our own conscious will. Trying to understand free will as though it were a kind of psychological causal process leads only to a mechanism that has no relationship at all to the experience of free will that we each have every day.

People appreciate free will as a kind of personal power, an ability to do what they want to do. Voltaire (1694–1778) expressed this intuition in saying, “Liberty then is only and can be only the power to do what one will” (1752/1924, p. 145). He argued that this feeling of freedom is not served at all by the imposition of randomness, asking, “would you have everything at the pleasure of a million blind caprices?” (p. 144). The experience of will comes from having our actions follow our wishes, not from being able to do things that do not follow from anything. And, of course, we do not cause our wishes. The things we want to do come into our heads. Again quoting Voltaire, “Now you receive all your ideas; therefore you receive your wish, you wish therefore necessarily. . . . The will, therefore, is not a faculty that one can call free. The free will is an expression absolutely devoid of sense, and what the scholastics have called will of indifference, that is to say willing without cause, is a chimera unworthy of being combated” (p. 143). A Free Willer, in short, would not generate the experience of conscious will.

We are left, then, with a major void. In leaving out a mechanism that might act like free will, theories have also largely ignored the experience of free will. The feeling of doing is a profoundly regular and important human experience, however, and one that each of us gets enough times in a day to convince us that we are doing things (non-randomly) much of the time. This deep intuitive feeling of conscious will is something that no amount of philosophical argument or research about psychological mechanisms can possibly dispel. Even though this experience is not an adequate theory of behavior causation, it needs to be acknowledged as an important characteristic of what it is like to be human. People feel will, and scientific psychology needs to know why. Clearly, people do not feel will because they somehow immediately know their own causal influence as it happens. The experience is the endpoint of the very elaborate inference system underlying apparent mental causation, and the question becomes: Why do we have this feeling?

3.1.2. Authorship emotion. Perhaps we have conscious will because it helps us to appreciate and remember what we are doing. The experience of will marks our actions for us. It helps us to know the difference between a light that we have turned on at the switch and a light that has flickered alive without our influence. To label events as our personal actions, conscious will must be an experience that is similar to an emotion. It is a feeling of doing. Unlike a cold thought or rational calculation of the mind alone, will somehow happens both in body and in mind. The experience of willing an action has an embodied quality, a kind of weight or bottom, which does not come with thoughts in general. In the same sense that laughter reminds us that our bodies are
Wegner: Précis of The illusion of conscious will

having fun, or that trembling alerts us that our bodies are afraid, the experience of will reminds us that we are doing something. Will, then, serves to accentuate and anchor an action in the body. This makes the action our own far more intensely than could a thought alone. Unlike simply saying “this act is mine,” the occurrence of conscious will brands the act deeply, associating the act with self through feeling, and so renders the act one’s own in a personal and memorable way. Will is a kind of authorship emotion.

The idea that volition is an emotion is not new. In fact, T. H. Huxley (1910) made the equation explicit: “Volition... is an emotion indicative of physical changes, not a cause of such changes. ... The soul stands to the body as the bell of a clock to the works, and consciousness answers to the sound which the bell gives out when struck. ... We are conscious automata.” Will is a feeling, not unlike happiness or sadness or anger or anxiety or disgust. Admittedly, conscious will does not have a standard facial expression associated with it, as do most other basic emotions. The look of determination or a set brow that is sometimes used to caricature willfulness is probably not identifiable enough to qualify as a truly communicative gesture. Still, will has other characteristics of emotion, including an experiential component (how it feels), a cognitive component (what it means and the thoughts it brings to mind), and a physiological component (how the body responds). Although conscious will is not a classic emotion that people would immediately recognize when asked to think of an emotion, it has much in common with the emotions.

The experience of consciously willing an action belongs to the class of cognitive feelings described by Gerald Clore (1992). He points out that there are a set of experiences such as the feeling of knowing, the feeling of familiarity, or even the feeling of confusion, that serve as indicators of mental processes or states, and that thus inform us about the status of our own mental systems. The experience of willing an action is likewise an informative feeling, a perception of a state of the mind and body that has a unique character. Although the proper experiments have not yet been done to test this, it seems likely that people could discriminate the feeling of doing from other feelings, knowing by the sheer quality of the experience just what has happened. The experience of willing is more than a perception of something outside oneself, it is an experience of one’s own mind and body in action.

Conscious will is the emotion of authorship, a somatic marker (Damasio 1994) that authenticates the action’s owner as the self. With the feeling of doing an act, we get a conscious sensation of will attached to the action. Often, this marker is quite correct. In many cases, we have intentions that preview our action, and we draw causal inferences linking our thoughts and actions in ways that track quite well our own psychological processes. Our experiences of will, in other words, often do correspond correctly with the empirical will, the actual causal connection between our thought and action. The experience of will then serves to mark in the moment and in memory the actions that have been singled out in this way. We know them as ours, as authored by us, because we have felt ourselves doing them. This helps us to tell the difference between things we are doing and all the other things that are happening in and around us. In the melee of actions that occur in daily life, and in the social interaction of self with others, this body-based signature is a highly useful tool. We resonate with what we do, whereas we only notice what otherwise happens or what others have done – so we can keep track of our own contributions without pencils or tally sheets.

Conscious will is particularly useful, then, as a guide to our selves. It tells us what events around us seem to be attributable to our authorship. This allows us to develop a sense of who we are and are not. It also allows us to set aside our achievements from the things that we cannot do. And perhaps most important for the sake of the operation of society, the sense of conscious will also allows us to maintain the sense of responsibility for our actions that serves as a basis for morality.

We can feel moral emotions inappropriately, of course, because our experience of conscious will in any given case may be wrong. The guilt we feel for breaking mother’s back may accrue via the nonsensical theory that we were culpable for her injury as a result of stepping on a crack. More realistically, we can develop guilty feelings about all sorts of harms we merely imagine before they occur – simply because our apparent mental causation detector can be fooled by our wishes and guesses into concluding that we consciously willed events that only through serendipity have followed our thoughts about them. By the same token, the pride we feel in helping the poor may come from the notion that we had a compassionate thought about them before making our food donation, although we actually were just trying to clear out the old cans in the cupboard. But however we do calculate our complicity in moral actions, we then experience the emotional consequences and build up views of ourselves as certain kinds of moral individuals as a result. We come to think we are good or bad on the basis of our authorship emotion. Ultimately, our experience of conscious will may have more influence on our moral lives than does the actual truth of our behavior causation.

3.2. How things seem

Sometimes how things seem is more important than what they are. This is true in theater, in art, in used car sales, in economics, and, it now turns out, in the scientific analysis of conscious will as well. The fact is, it seems to each of us that we have conscious will. It seems we have selves. It seems we have minds. It seems we are agents. It seems we cause what we do. Although it is sobering and ultimately accurate to call all this an illusion, it is a mistake to conclude that the illusion is trivial. To the contrary, the illusions piled atop apparent mental causation are the building blocks of human psychology and social life. It is only with the feeling of conscious will that we can begin to solve the problems of knowing who we are as individuals, of discerning what we can and cannot do, and of judging ourselves morally right or wrong for what we have done.

Usually, we assume that how things seem is how they are. We experience willing a walk in the park, winding a clock, or smiling at someone, and the feeling keeps our notion of ourselves as persons intact. Our sense of being a conscious agent who does things comes at a cost of being technically wrong all the time. The feeling of doing is how it seems, not what it is – but that is as it should be.

3.3. Postscript

This précis of The Illusion of Conscious Will is an abridgment of three of the book’s chapters. It focuses on the main
Commentary/Wegner: Précis of The illusion of conscious will

Open Peer Commentary

The self is virtual, the will is not illusory

George Ainslie
Veterans Affairs Medical Center, Coatesville, PA 19320.
george.ainslie@med.va.gov www.picoeconomics.com

Abstract: Wegner makes an excellent case that our sense of ownership of our actions depends on multiple factors, to such an extent that it could be called virtual or even illusory. However, two other core functions of will are initiation of movement and maintenance of resolution, which depend on our accurate monitoring of them. This book shows that will is not an imponderable black box but, rather, an increasingly accessible set of specific functions.

This book is an encyclopedic analysis of the ways in which our sense of volition fools us. Wegner (2002) has assembled a remarkably broad range of examples wherein people behave without being aware of deciding to do so; falsely believe that they are deciding; or, most subtly, experience a decision as occurring at a different time than objective evidence places the decision. I think that Wegner over-reads the implications of these examples when he calls conscious will an illusion. Our eyes sometimes fool us, too, as when we mislocate an underwater object or are led by contextual cues to misjudge the size or distance of an object, but we still say that we are actually seeing it. The famous moon illusion does not make the moon illusory. Wegner has many valuable things to say, but the examples he assembles to argue against conscious will apply to only parts of what his own material demonstrates to be a complex phenomenon. I submit that what he—and we—call conscious will comprises at least three somewhat independent processes, two of which depend on the person’s accurate sense of their operation.

Dealing with these two first: The initiation of movement and the maintenance of resolution, perhaps Wegner’s “little dabs” of will and its “long lasting property,” respectively, each has its kind of proprioception within the mind (brain?) itself; we rely on the accuracy of this proprioception from minute to minute, day in and...
Commentary/Wegner: Précis of The illusion of conscious will
day out. One of Wegner's own examples illustrates the *initiation of movement* part of the will to move. The amputee who is conscious of moving nonexistent toes is obviously not relying on peripheral sensations. She reports mentally doing what, in someone with toes, accurately governs their movements. By abnormally removing the peripheral component of this process, nature has isolated Hume's "impression of the feel and consciousness of the will. Whenever knowingly give rise to any new motion of our body" (Hume 1739/1888, p. 399, quoted in sect. 1.1.1 of the target article, emphasis in original). The associated movements are gone, but the experience of will in this trivial sense of connecting mind and body remains, and there is no reason to believe that the subject's consciousness of its operation *per se* is inaccurate, despite the illusory downstream effects. This consciousness is different in kind from mere association of a tree branch actually moved without my prior appreciation of will every time I thought of its moving, it would not feel as if I suddenly had a previously unrecognized muscle, but instead would probably give me the eerie sense of having my mind read (see Gray Walter's experiment in Dennett 2003a, p. 240).

*Maintenance of resolution* is more important. It is where both stages of the conception of will reside, and our beliefs about it have practical effects on self- and social control. Defending direct perception of this resolution is hard because, although observers have agreed on many functional properties—the effects of practice, of reference to principles, of single lapses, and so forth (Ainslie 2001, pp. 119–20)—they have not agreed on a way of describing the thing itself. I have argued that resolution is not a thing, or unitary sensation, at all, but an intertemporal process analogous to bargaining, and that it is just as directly reportable as the events of interpersonal bargaining are (Ainslie 2001, pp. 90–104). Briefly: The way we make our intentions consistent is to perceive our current decision as a test case for how we will decide similar choices generally, so that our expected reward from consistent intention is staked on "cooperating" with our future selves and is sharply reduced if we "defect" to an impulsive alternative. Although people sometimes make mistakes of this conscious priority, they are not as prone to oversight as to misestimation of their morality, principle, personal intention, and even divine help, we universally experience a big stake as resolve and a lapse as a loss of part of this stake, engaging guilt. The proprioception here is the recursive self-monitoring process, the testing of our will, which is not prominent in behaviors we are confident of executing but is glaringly evident when we resolve to resist a favorite vice or to dive into a cold lake. The mind's compass to which Wegner sometimes considers the experience of will to be "a feeling" (Wegner 2002, p. 3), directly sensed, "not unlike happiness or sadness" (p. 326). However, he more often considers it a "fabrication" (p. 3), a cognitive construct combining what he calls priority, consistency, and exclusivity (p. 69). For some of us, the idea that will is a directly-sensed feeling suggests a search for the neural correlates of this feeling (Bogen 1997). Wegner briefly refers to stimulation of the exposed cortex by Penfield and the well-con-}

**The experience of will: Affective or cognitive?**

Joseph E. Bogen
Neurologic Surgery, University of Southern California, Los Angeles, and Division of Biology, California Institute of Technology, Pasadena, CA 91105.

**Abstract:** Wegner vacillates between considering the experience of will as a directly-sensed feeling and as a cognitive construct. Most of his book is devoted to examples of erroneous cognition. The brain basis of will as an immediately-sensed emotion receives minimal attention.

Wegner sometimes considers the experience of will to be "a feeling" (Wegner 2002, p. 3), directly sensed, "not unlike happiness or sadness" (p. 326). However, he more often considers it a "fabrication" (p. 3), a cognitive construct combining what he calls priority, consistency, and exclusivity (p. 69). For some of us, the idea that will is a directly-sensed feeling suggests a search for the neural correlates of this feeling (Bogen 1997). Wegner briefly refers to stimulation of the exposed cortex by Penfield and the well-con-
We believe in freedom of the will so that we can learn

Clark Glymour
Department of Philosophy, Carnegie Mellon University, Pittsburgh, PA 15213, and Institute for Human and Machine Cognition, University of West Florida, Pensacola, FL 32507. eg09@andrew.cmu.edu

Abstract: The central theoretical issue of Wegner’s book is: Why do we have the illusion of conscious will? I suggest that learning requires belief in the autonomy of action.

You should believe in freedom of the will because if you have it you’re right, and if you don’t have it you couldn’t have done otherwise anyway.

—Sam Buss (Lecture at University of California, San Diego, 2000)

Wegner’s (2002) fascinating book argues that conscious will is like the existence of God: most everyone believes it most of the time, but it isn’t so. (The simile is mine, not Wegner’s.) Hence, what I take to be the central theoretical issue of the book: Why do we have the illusion of conscious will so systematically and so persistently? Perceptual illusions are explicable as unusual violations of the conditions under which our sensory processing are veridical,

Calling in the Cartesian loans

Daniel C. Dennett
Center for Cognitive Studies, Tufts University, Medford, MA 02155. d dennett@tufts.edu http://ase.tufts.edu/cogstud/~dennett.htm

Abstract: Wegner’s tactic of describing the conscious mind as if it inhabited a Cartesian Theater in the brain is a stopgap solution that needs to be redeemed by paying off these loans of comprehension. Just how does Wegner propose to recast his points?

Three quotations from Wegner’s (2002) book, each not just defensible but, I think, importantly insightful, take out Cartesian loans that are now overdue.

“We can’t possibly know (let alone keep track of) the tremendous number of mechanical influences on our behavior because we inhabit an extraordinarily complicated machine” (p. 27). These machines “we inhabit” simplify things for our benefit. Who or what is this “we” that inhabits the brain? A Cartesian ghost in the machine? Surely not, in spite of first appearances.

“Conscious will is particularly useful, then, as a guide to ourselves” (p. 326). Again, who or what uses this handy guide? Does one part of the brain use another part? Is it as simple as that?

“Illusory or not, conscious will is the person’s guide to his or her own moral responsibility for action” (p. 341). My body is causally responsible for whatever effects emanate from it, whether it is falling down a flight of stairs, or pulling the trigger of a gun, but I, the person “inhabiting” this body, am morally responsible only for my actions. Again, who is this person and what is he doing in my body?

I have defended Wegner’s tactic of temporarily indulging in these ways of speaking, and sketched a way for him to recast his points without relying on the ominous image of a Cartesian Theater in which the Self sits as Witness and Decision-Maker (Dennett 2003a; 2003b; 2003c). But I would like to see how he himself proposes to pay off these comprehension-loans, since he may have some other tricks up his sleeve.
but attributions of free will are scarcely unusual, and an explanation is required. It is hard to resist attributing autonomy to others, even when we see the mechanics of reason come apart before our eyes. Anyone who has had day-to-day encounters with someone suffering from obsessive/compulsive disorder will have had the impulse to blame the sufferer for irrational actions committed in the course of their otherwise normal conduct and discourse. If we have no Cartesian freedom of the will, why do we have so fierce an inclination to attribute autonomy to ourselves and others? What function, what cognitive causal role, do such beliefs have that might help to explain their emergence and retention in the human psyche, and why do we have them consciously? Wegner offers an answer to the first of these twinned questions. I will offer another.

The illusion of conscious will

Consider scientific inference from observational, non-experimental, data. There are several possible explanations for a correlation observed among two kinds of events for which instances of one kind precede those of the other: Events of the first kind may cause the second; or some third factor or factors may influence both kinds of events; and there are still other possibilities. For concreteness, consider an association between smoking and lung disease, which could be explained by at least two different causal structures:

1. Smoking → Lung Disease
2. Smoking → Unknown → Lung Disease

To make a reasonable causal inference, one must have grounds to exclude confound explanation, one must do, and that is why observational science is hard. Experimentation tends to eliminate alternative explanations of data. What makes an experiment an experiment is that acting from outside the system under study, the experimenter determines the value of the causal variable, or determines its probability distribution. If the experimenter fixes or randomizes the value of the causal variable in each case, and does so by a method not influenced by other features not under the experimenter's control, then there is no confounding. If we force someone—or an entire population—not to smoke, then we eliminate confounding, and, if smoking does not cause lung disease then these two variables are uncorrelated in the experimental results. (For mathematical details, see Pearl 2000; Spirtes et al. 2001; and for a philosophical exposition, see Woodward 2003.)

Independent manipulation does not make causal learning possible, but it makes it enormously easier to make accurate causal inferences. Whatever the circumstances, if one does not impose the premise—warranted or not—that the association of putative cause and effect is not produced by other common causes of both, the inference to causation is wanton.

For our inner workings—the unconscious, biological algorithms of thought—to allow that actions have unknown causes would be precisely for them to allow that those unknowns might also cause the immediate and slightly more remote events that we take to be effects of actions; action and event would be potentially confounded and no causal inference would be possible in everyday life, just as no causal conclusions are possible in ill-designed, confounded, scientific experiments or in poorly designed observational studies. So, unconsciously at least, to be intelligent in the way we are, we must presuppose autonomous actions—and to make correct causal inferences, actions and their effects must for the most part actually be unconfounded by common causes. An organism that did not so assume might learn by association, but its ability to plan and foresee the effects of interventions in the world would be severely limited. Daniel Povinelli (2000) and Tomasello and Call (1997) give evidence that our nearest biological neighbors are limited in these respects, while Gopnik et al. (2004) give evidence that even quite young children make comparatively sophisticated causal inferences from data in which passive correlations and effects of interventions are combined. If, from whatever causes, the assumptions of our inner processes that lead to action are consciously manifested in the very instance of action or in the perception of action in others, we will have the conscious sense of autonomous agency, of freedom of the will. And we do. We think immediately that our actions cause the observed effects, and nothing else causes both our actions and the observed effects. Usually, we assume the same of others, and if we did not then we could not learn causal relations from their actions and the events that follow them.

The elusive illusion of sensation

Valerie Gray Hardcastle

Department of Science and Technology in Society, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061-0247. valerie@vt.edu

http://www.mind.vt.edu/

Abstract: The sensation of will is not the same thing as the will itself any more than the sensation of hunger is the same thing as being devoid of nutrients. This is not a really surprising claim, but it is the only claim to which Wegner is entitled in his book.

When I feel hunger pangs, am I feeling genuine hunger, or am I feeling "merely" the sensation that accompanies real hunger, a purely physiological state? If the latter, then hunger pangs must be some sort of illusion, a stand-in for states we cannot access consciously. When our bodies infer that they need more nutrients, we feel hungry. However, as the popular press makes very clear, we are often wrong about this inference and consequently feel hungry when we aren't really.

This mediation on hunger parallels what Wegner (2002) says about our sensations of willing an action. The sensation of willing isn't actually doing anything; it certainly isn't causing our bodies to behave in any particular way. Instead, the sensation is "merely" telling us that (we think) our own psychological states are driving our bodies.

Is this conclusion so surprising? I grant that we generally talk and think about the will in very sloppy terms, but when we get right down to it, do we really believe that the sensation of willing just is the will itself? I submit that we do; we believe, if we have even thought about these matters before, that the sensation informs us about the sort of actions we are performing. If we feel the force of our will, then we believe that we, in some important and fundamental sense, are the causal agents responsible for what we are doing. The sensation of will isn't the will itself any more than the sensation of hunger is the same thing as being devoid of nutrients, or the sensation of warmth is heat itself, or the smell of a rose is the rose itself. In each case, our sensations tell us something about the world out there (or in here); they indicate or represent to us the way the world is (or take it to be).

Wegner provides us with case after case of how our sensations...
of will are mistaken, how we sometimes do things ourselves but attribute these actions to others, how we sometimes think we are doing things ourselves, but we aren't. He is right; our sense of will is sometimes – maybe a lot of the time – misleading.

But so what? What, if anything, does this tell us about freedom of the will – the actual will, not what we sense as a marker for the will? To much. In order to know something about the actual will for example, whether it exists in any interesting sense, we would have to know how the sensation of will connects up with either our underlying psychology or our underlying physiology or both. However, unlike the case of hunger, in which we know a lot about the connection between various levels of hormones in our blood stream and wanting to eat, we know very little about what the sensation of will actually reflects. Maybe it does mirror a genuine self in the brain, our central control that initiates or at least approves our purposeful behaviors. Maybe it doesn't. But knowing that our conscious sensations of will are sometimes mistaken doesn't shed any light on this topic.

We know some actions happen to us – I sneezed in the middle of lecture – and others have a psychological reason behind them – I raised my hand in the middle of lecture. We can tell the difference between these sorts of activities, both from the inside, as it were, and from the outside. But what is this difference? Is it just that the latter is accompanied by a sensation of will and the former isn't? Is it just that we explain the latter in terms of beliefs and desires and the former in terms of physiology? Or does the latter occur as the endpoint in a causal chain mediated by my own psychological states, whereas the former doesn't? I think that no matter what one's metaphysical stripe, one would have to agree with the last suggestion: What differentiates willed actions from actions that are not willed is the causal history of the action. Willed actions flow from or through my psychological states in ways that unwilled actions don't.

But if this is the case, then in what sense is our sensation of conscious will an illusion? Our sensation serves to differentiate which actions flowed from or through our psychological economy from those that didn't. We must get to know in what way we make it wrong lots of times. Nevertheless, the sensation is reflecting something real, as real as our bodies' need for nutrients. The important question is what exactly is that sensation reflecting.

Wegner wants to argue that we don't really have selves, that our sensation of selfhood, too, is just another inference our bodies and brains make about what we are doing things ourselves, but we aren't. He is right; our sense of will is sometimes – maybe a lot of the time – misleading.

When an object, say a boat on the water, moves away, its retinal image decreases in size. However, instead of experiencing the boat as shrinking, the viewer experiences it as receding into the distance. This could be called an illusion; the retinal image is getting smaller, not further away. However, to say that this is an “illusion” is to ignore the determinants of object constancy. When the viewer's understanding of boats and the three-dimensional world are included in the analysis of visual experience, the correlation between a shrinking retinal image and the perception of a constant-sized but increasingly distant object is perfectly understandable. Or, to put it another way, to say that one of the visual constancies is an “illusion” is to overlook that there is more to vision than the retina.

Wegner's treatment of conscious will (Wegner 2002) is rather like trying to account for object constancy while limiting the analysis to the retina. He emphasizes that conscious will is an inference and that its contents often do not match up well with the actual factors that cause voluntary action. For instance, we may be aware of the intention to raise our hand (or assume this intention after the fact), but not be aware of the determinants of this intention or of having made an inference. From these “discrepancies,” Wegner concludes that conscious will is an illusion. However, as in the object constancy example, a more complete account of the input eliminates the illusion.

Object constancy is about the fact that we live in a three-dimensional world and that when objects move, they usually do not change shape. What is conscious will about? What is its stimulus? The answer cannot be found in a textbook (as with the perceptual constancies), but it is familiar and easily identified.

As documented by Wegner, conscious will's domain is behavior, in particular our own behavior. Just as perception tracks dimensions of the external world, conscious will tracks the important fact that our own activities vary in the degree to which they are influenced by consequences (e.g., rewards, incentives, punishments, and the like), by the values we adopt, and by new information. Some activities are immune to these factors, whereas others are easily modified by just a hint of praise or disapproval. For instance, consider the different causal relations relating to a patellar reflex and learning to kick a ball and flushing the toilet or applying rouge, a defensive blink and a conspiratorial wink at a friend. The second activity in each comparison we call voluntary, and the first we call involuntary. The distinction is not a matter of free will versus determinism. Antecedents govern voluntary and involuntary acts. Rather, the mediating neural architecture and nature of the antecedents differ. Differences in neural connections allow for variation in the degree to which activities are influenced by exogenous or endogenous causes. The point is, conscious actions do not depend on intentions or other subjective reports. We can be conscious of involuntary acts (I know I am going to blink, but I can't help it), and as Wegner's literature review ably demonstrates, we can be unaware of voluntary acts. In other words, voluntariness (susceptibility to consequences) provides an objective basis for subjective experience, just as the conservation of an object's shape and size while moving provides a basis for perceptual constancy.

Wegner acknowledges that behavior varies with regard to its susceptibility to consequences (e.g., the ear wiggling discussion, Wegner 2002, pp. 31–34), and also acknowledges that voluntary actions are the usual focus of conscious will. However, these observations are made in passing, and his analysis proceeds without any further discussion of the objective basis for the sensation of doing something.” Given this omission it is understandable that he concludes that it is an illusion. This is not to say that conscious will is a literal reflection of susceptibility to rewards. For instance as Wegner's discussion of automatic processes (2002, pp. 56–59) demonstrates, many learned, reinforced actions can move out of awareness.

Leaving out the objective correlates for conscious will leads to empirical and logical problems. An empirical shortcoming is the de-emphasis of the contribution that conscious will makes to voluntary action. Often Wegner seems to be saying that conscious will is no more than an after-the-fact frill, at best useful for a kind of moral bookkeeping (see below). I am not sure that this is what he

The sense of conscious will

Gene M. Heyman
BPRL and Psychiatry, McLean Hospital and Harvard Medical School, Belmont, MA 02478. gheyman@mclane.harvard.edu

Abstract: Wegner's conclusion that conscious will is an illusion follows from a key omission in his analysis. Although he describes conscious will as an experience, akin to one of the senses, he omits its objective correlate. The degree to which behavior can be influenced by its consequences (voluntariness) provides an objective correlate for conscious will. With conscious will anchored to voluntariness, the illusion disappears.

Commentary/Wegner: Précis of The illusion of conscious will

Commentary/Wegner: Précis of The illusion of conscious will

Commentary/Wegner: Précis of The illusion of conscious will

Commentary/Wegner: Précis of The illusion of conscious will
Commentary/Wegner: Précis of The illusion of conscious will

really means, because it is easy to show that the contents of consciousness (e.g., plans) can alter the course of voluntary action. To make a less obvious point: Voluntary behavior is subject to competing contingencies, and without conscious awareness of the more global ones (those that are good for us in the long run), we would always fall victim to the most immediate reward (e.g., Heyman 1996, 2003). This observation is celebrated in Greek myth (e.g., the story of Odysseus and the sirens) and is embedded in moral and spiritual teachings (which can be seen in part as pleas for attending to vital but less salient, long-term contingencies).

The logical problem is that leaving out the stimulus leads to circular accounts of how conscious will arises and its purpose. According to Wegner, associative correlations are sufficient for the sense of conscious will. He writes that if intentions or plans are followed by action, the mind infers that the action was caused by conscious will. However, intentions and plans imply the sensation of conscious will. In order to have an intention or plan regarding action, one must already have the belief that behaviors exist that can be modified by goals and consequences. By leaving out the behavioral basis for plans (voluntariness), his account amounts to the circular statement that conscious will (intentions and plans) is the basis of conscious will.

There is a similar logical problem with Wegner’s theory of what conscious will is good for—the purposes it serves. He asks (p. 325): Why do we have the feeling of conscious will? His general conclusion is that it serves as a guide for moral responsibility. His argument is that we should only be responsible for the actions that we intended, and, hence, that conscious will serves the purposes of the moral order. However, the vocabulary of moral interactions assumes “intentions” and “free choices.” The view is supported by several lines of evidence in which conscious will is dissociated from the actual performance of voluntary movements, contrary to the first view. With the first view discarded, how can we explain the contradictory situation where we feel we are willfully causing an action that is in fact a product of a certain mechanism (such as the fear of punishment). Hence, moral responsibility (as defined by Wegner) cannot explain why we have the feeling of conscious will; it builds on its prior existence. The way out of this circularity is to identify the objective correlates of the sensation of will.

Wegner’s literature review and his own experiments make it clear that conscious will is in many respects like the basic senses. Like the senses it is correlated with an important dimension of the objective world; like the senses it does not provide a literal representation of either the objective world or the proximal stimulus; and like the senses it has proven a useful guide to more effective voluntary actions. My title for the experimental literature Wegner reviews would have been: The sense of conscious will.

ACKNOWLEDGMENT
I thank Martha Pott for her helpful comments.

How neuroscience accounts for the illusion of conscious will

Masao Ito
Brain Science Institute, Institute RIKEN, Wako, Saitama, 351-0198, Japan.
masao@brain.riken.jp

Abstract: Wegner’s monograph presents the view that conscious will is a feeling that we experience when we perform an action through a mechanistic process of the brain, rather than a mental force that causes the action. The view is supported by several lines of evidence in which conscious will is dissociated from the actual performance of voluntary movements, as in automatism. The book further extends an insightful analysis of the mechanisms underlying consciousness, in which the cerebral cortex is generally involved, a learned movement can then be performed unconscious through the cerebellar pathway. This condition closely resembles that illustrated in Figure 3.1 (Wegner 2002, p. 68) for explaining conscious and unconscious mental events behind voluntary movement.

The internal model hypothesis has been expanded to problems of the thought (Ito 1993). When we think, the prefrontal cortex acts as an executive cortex and manipulates an image, a concept, or an idea, which are collectively termed the mental model. A mental model is formed by combining various pieces of information received from the sensory cortex, and is stored in the temporal parietal cortex that constitutes the internal environment of the brain. Just as we manipulate an arm or a leg during movements, we manipulate a mental model during thoughts. During repeated trials of thought, a mental model in the temporoparietal cortex is copied in a cerebellar internal model. By referring to such a copy of a mental model, the thought can be performed quickly and unconsciously of its processes. This fits the situation described on page 67: “when you multiply 3 times 6 in your head, the answer just pops into your mind without any indication of how you did that” (Wegner 2002). Such a calculation, when first performed in the cerebral cortex, must require conscious effort, but as a learnt calculation it is performed in the cerebellum, and will no longer rise to the level of consciousness.
Meditation, St Louis, MO 63110; probbings@artsci.wustl.edu
http://www.nil.wustl.edu/labs/corbetta/personnel/ajack.html
http://artsci.wustl.edu/~probblings/home

Abstract: Wegner’s thesis that the experience of will is an illusion is not just wrong, it is an impediment to progress in psychology. We discuss two readings of Wegner’s thesis and find that neither can motivate his larger conclusion. One sees science requiring us to dismiss our experiences, its real promise is to help us to make better sense of them.

Dan Wegner has written a book with an awful lot to recommend it. This amusing, engaging, and intelligent text reviews fascinating phenomena and offers theoretically ingenious explanations of them. Nonetheless, we are here to talk about the problems. The greatest problem the reader faces is that of answering a straightforward question: What is Wegner’s central thesis? We couldn’t find one consistent answer, so we will discuss the two most plausible interpretations of “the illusion of conscious will.” We will show that neither can support Wegner’s view that he has radically undermined our concept of self, the view boldly stated in a concluding paragraph as follows: “it seems to each of us that we have conscious will. It seems to us that we have minds. It seems we are agents. It seems we cause what we do... it is sobering and ultimately accurate to call all this an illusion” (Wegner 2002, pp. 341–42). Contrary to Wegner’s conclusion, in which mechanistic explanation triumphs and mentalistic explanation is dismissed as illusion, we will argue that mechanistic understanding can actually enrich our concept of mind.

We consider two readings of Wegner’s thesis, one metaphysical and the other epistemological. The metaphysical reading is as follows: Though it appears from the first-person perspective that our actions are caused by our conscious intentions to act, the real springs of action lie elsewhere. Appearances (and folk wisdom) notwithstanding, conscious intentions are never implicated in the production of behavior. However, it may be pointed out that mental models and their cerebellar copies, if involved in high levels of mentations, would give rise to a conception of such a causal agency that predicts and instructs our behaviors.

The illusionary triumph of machine over mind: Wegner’s eliminativism and the real promise of psychology

Anthony L. Jacka and Philip Robbins

*Department of Neurology, Washington University in St Louis School of Medicine, St Louis, MO 63110; Department of Philosophy, Philosophy-Neuroscience-Psychology Program, Washington University in St Louis, St Louis, MO 63130-4899: ajack@npg.wustl.edu
probbings@artsci.wustl.edu

http://www.nil.wustl.edu/labs/corbetta/personnel/ajack.html
http://artsci.wustl.edu/~probblings/home

Causal agency. The mistake in interpreting the experience of will as an actual causal mechanism may arise, as this book explains, from our compelling conception of causal agents. When we perceive our mind, we have an agent that does things purposefully (Wegner 2002, Chs. 5, 7). The agent develops as a child grows, and governs our mental life. It may culminate to an idea of God or a soul-like being. All this is a fabrication, nothing but a way of making sense of behavior. However, it may be pointed out that mental models and their cerebellar copies, if involved in high levels of mentations, would give rise to a conception of such a causal agency that predicts and instructs our behaviors.

Causal Transparency

...
Commentary/Wegner: Précis of The illusion of conscious will

there is for us to doubt the reality of the external world that we perceive around us.\(^5\)

Wegner fails to realize that we are not forced to interpret our experience of agency in just one way. We can make sense of our experience in many different ways, just as we can perceive events in the external world in different ways, depending on our implicit or explicit theoretical assumptions. Effectively, Wegner supposes that our concept of mind must remain frozen in a naïve folk-psychological model. By doing so, he is failing to realize the true promise of psychology: that psychological research can have a major impact by improving upon folk psychology. In Wegner's worldview, the scientific project inevitably reduces us to mindless mechanisms. In contrast, we believe science's greatest achievement will be that of transforming our personal and cultural understanding of ourselves to better correspond with human nature.

NOTES

1. To be fair to Wegner, this theoretical distinction has only recently come to the fore (Jack & Shallice 2001; Schooler 2002; Lashme & Marcel 2002).

2. Wegner explains automatisms via his theory of ironic processes. The idea is that the conscious intention not to perform a certain action actually has the effect of giving rise to the action that the subject is trying to inhibit. Wegner has produced substantial evidence that inhibitory mental sets have such ironic effects in other contexts (notably thought suppression). Although the conscious intention causes the action, the subject does not experience the action as willed because the action is inconsistent with the aim of the intention.


4. The belief is so abstract that it is hard to imagine what it would be like to have it. Direct access implies certain knowledge, so the illusion would cause the subject to believe that their experience of agency cannot be mistaken. If you can doubt your experience of agency, then you cannot be suffering from an illusion of causal transparency.

5. The argument that Wegner implicitly relies on to reach his profoundly skeptical view of the mind closely echoes the argument Descartes used to derive his skepticism about the external world. In both arguments the demonstration that we can be mistaken on occasion is used to motivate the much more radical view that we should question everything. The demonstration that we can be mistaken, but Wegner's book appears to be a reassertion of the automaton-theory in pure form. His very first chapter argues that “It usually seems that we consciously will our voluntary actions, but this is an illusion” (Wegner 2002, p. 1). Just to make his point clear, Wegner offers (Fig. 3.1, p. 68) a diagram showing an “actual causal path” between an unconscious cause of action and conscious action, and another “actual causal path” between an unconscious cause of thought and conscious thought, but only an “apparent causal path” (italics in original)—the experience of consciousness will be between conscious thought and conscious action. He concludes with Albert Einstein’s image of a self-conscious but deluded moon, blithely convinced that it is moving of its own accord. In Wegner’s view, apparently, we are conscious automata after all.

Wegner musters a great deal of evidence to support his claim that our experiences of voluntary and involuntary action are illusory, including an entire chapter devoted to hypnosis. In fact, Wegner goes so far as to note that “hypnosis has been implicated in a conscious capacity, and are strictly unconscious. By contrast, controlled processes lack these properties, and are—although many scientific psychologists do not like to use the term—reflections of ‘conscious will.’

To many of us, this seems to be a perfectly reasonable compromise, but Wegner’s book appears to be a reassertion of the automaton-theory in pure form. His very first chapter argues that “It usually seems that we consciously will our voluntary actions, but this is an illusion” (Wegner 2002, p. 1). Just to make his point clear, Wegner offers (Fig. 3.1, p. 68) a diagram showing an “actual causal path” between an unconscious cause of action and conscious action, and another “actual causal path” between an unconscious cause of thought and conscious thought, but only an “apparent causal path” (italics in original)—the experience of consciousness will be between conscious thought and conscious action. He concludes with Albert Einstein’s image of a self-conscious but deluded moon, blithely convinced that it is moving of its own accord. In Wegner’s view, apparently, we are conscious automata after all.

Wegner musters a great deal of evidence to support his claim that our experiences of voluntary and involuntary action are illusory, including an entire chapter devoted to hypnosis. In fact, Wegner goes so far as to note that “hypnosis has been implicated in a conscious capacity, and are strictly unconscious. By contrast, controlled processes lack these properties, and are—although many scientific psychologists do not like to use the term—reflections of ‘conscious will.’

James was clearly committed to a causal role for consciousness, and thus for free will, but his statement implied a willingness to alter his view, when warranted, as psychology advanced. Indeed, the behaviorist revolution carried with it a resurgence of the automaton theory, reflected in Watson’s emphasis on conditioned reflexes and Skinner’s emphasis on stimulus control (Tolman’s purposivist interpretation of learning was an exception). On the other hand, the cognitive revolution implied an acceptance of James’ functionalist view: the primary reason to be interested in beliefs, expectations, and mental representations is that they have some causal impact on what we do. In fact, modern cognitive psychology accepts a distinction between automatic and controlled mental processes (e.g., Logan 1987; Shiffrin & Schneider 1984): Automatic processes are inevitably evoked following the presentation of some cue, are incorrigibly executed, consumed in a fraction of a second, and performed using prepotent responses. In contrast, controlled processes require effort and can be deliberately interfered with. The question of whether mental processes are automatic or controlled is at the heart of modern cognitive psychology, and it is clear that Wegner’s book will have an impact on the field.

“An unwarrantable impertinence”

John F. Kihlstrom

Department of Psychology, University of California, Berkeley, Berkeley, CA
94720-1650. kihlstrm@socrates.berkeley.edu
http://socrates.berkeley.edu/~kihlstrm

Abstract: Wegner’s many examples of illusory involuntariness do not warrant the conclusion that the experience of voluntariness is also an illusion. His arguments appear to be related to the contemporary emphasis on automaticity in social cognition and behavior; both appear to represent a revival of situationism in social psychology.

In his *Meditations* of 1641, Descartes asserted that consciousness, including free will, sharply distinguished man from beast (cf. Descartes 1641/1680), and thus he initiated the modern philosophical and scientific study of the mind. As time passed, however, philosophers became more materialistic and began denying this distinction, most visibly Julien Offray de la Mettrie, whose *Man a Machine* (Mettrie 1748/1749) claimed that humans were conscious automata, and Shadworth Holloway Hodgson, whose *The Theory of Practice* (Hodgson 1870) introduced the term *epiphenomenalism*. Although materialist monism was highly attractive to those who would make a science of psychology, William James, in his *Principles of Psychology* (James 1890/1980, p. 141), dismissed “the automaton-theory” as “an unwarrantable impertinence in the present state of psychology” (emphasis in original).
Commentary/Wegner: Précis of The illusion of conscious will

Self-reports of intentionality... may be attributions or interpretations based on a priori, implicit theories of behavior and on perceptions of the stimulus situation... Experiences of volition and involuntariness... are constrictions or interpretations made possible by the high degree of automaticity that is characteristic of all complex behavior. (Kirsch & Lynn 1995)

Based on this thesis, we reached the conclusion that “behavior, including novel and intentional behavior, is initiated automatically” (Kirsch & Lynn 1999a, p. 504). Therefore, we are pleased to see such a thorough explication of this idea. Unfortunately, Wegner’s discussion of hypnosis is inconsistent and misleading. The aim of this review is to correct these errors.

The phenomena of hypnosis. Hypnosis consists of two components: an induction procedure (e.g., “you are becoming hypnotized”) and suggestions that are usually given after the induction (Wegner refers to these as “tests”). Up to 90% of the population respond to at least some hypnotic suggestions (Kirsch et al. 1995). Thus, hypnotic phenomena are very relevant to automaticity and the illusion of will. Wegner’s “cautionary note” (Wegner 2002, p. 255) notwithstanding.

In examining these hypnotic phenomena, Wegner overestimates the role of inducing hypnosis and underestimates the importance of suggestions. Hypnotic suggestibility scales are not “indications of the success of the induction” (p. 282). These scales assess participants’ responses to hypnotic suggestions. Usually, this is done after inducing hypnosis, leading Wegner to believe that the responses are indications of “unique abilities possessed by those who are hypnotized” (p. 293). However, these responses can also be elicited without a hypnotic induction. In response to suggestions, people experience automatic movements, inhibited movement, hallucinations, pain reduction, and suggested amnesia, all without the induction of hypnosis. The effect of a hypnotic induction is to increase responsiveness to these suggestions, but only to a surprisingly small degree (“far less than the classical hypnotists would have supposed had the question ever occurred to them,” wrote Clark Hull [1933, p. 298]) and only for a minority of subjects (Barber & Glass 1962; Braffman & Kirsch 1999; Hilgard & Tart 1966; Hull 1933; Spanos et al. 1985; Stan & Spanos 1980; Weitzenhoffer & Sjoberg 1961). Suggestion without hypnosis has even been found to reduce warts (DuBreuil & Spanos 1993) and control pain during surgery without anesthesia (Jones 1999).

Wegner also overestimates the degree of subjective automaticity in hypnosis, thereby reinforcing the mythology of hypnosis perpetuated in novels, movies, and stage presentations. He asserts that hypnosis involves a “giving over control to the hypnotist” (p. 271), in which “the subject may perceive a draining away of conscious control...” and only for a minority of subjects (Kirsch et al. 1995; Kirsch & Lynn 1999a, p. 504). Therefore, we are pleased to see such a thorough explication of this idea. Unfortunately, Wegner’s discussion of hypnosis is inconsistent and misleading. The aim of this review is to correct these errors.

The illusion of conscious will

Although there are a few dissenters (Kirsch & Lynn 1997; 1998a; 1998b; Woody & Bowers 1994; Woody & Sadler 1998), most theorists of hypnosis, whatever their other disagreements, agree that the experience of involuntariness in response to hypnotic suggestions is in some sense illusory. In Hilgard’s (1977) neodissociation theory of divided consciousness, the experience of involuntariness results from the subject’s lack of conscious awareness of the volitional activities required to execute the suggestion (see also Kilbourn 1992b). From a social-psychological perspective, Sarbin and Coe (1972) identified the description of hypnotic phenomena as “happenings” rather than “doings” as central to the hypnotic role. Similarly, Spans (1980a; 1980b; Spanos et al. 1985) characterized reports of involuntariness as a strategy for convincing others that one was really hypnotized, and identified some of the conditions under which subjects could actually persuade themselves that such reports were true.

In fact, most of the other phenomena described at length by Wegner, such as the Chevreul pendulum, automatic writing, the Ouja board, and even facilitated communication, have this quality: behavior that is experienced by the individual as involuntary is actually voluntary in nature. Documenting this illusion would make for an interesting book, as indeed it has (Spitz 1997). But Wegner puts this evidence to a different rhetorical use – he tries to convince us, by citing examples of illusion involuntary behavior, that our experience of voluntary behavior, in the ordinary course of everyday living, is illusory as well. Logically, of course, this does not follow. To be sure, there exist illusions of control as well (e.g., Alloy et al. 1989), but even these do not just the strong conclusion that all experiences of voluntariness are illusions – which is what Wegner seems to be claiming.

Given that the evidence for an illusion of voluntariness is weak, the rationale for Wegner’s claim must be found elsewhere – in theory, or perhaps in ideology. In this respect, Wegner’s book can be seen as an attempt to counteract the increasing trend in contemporary social psychology, that our experience of voluntary behavior, in the ordinary course of everyday living, is illusory as well. Logically, of course, this does not follow. To be sure, there exist illusions of control as well (e.g., Alloy et al. 1989), but even these do not just the strong conclusion that all experiences of voluntariness are illusions – which is what Wegner seems to be claiming.

Hypnosis and will

Irving Kirsch\(^a\) and Steven Jay Lynn\(^b\)

\(^a\)Faculty of Health and Social Work, University of Plymouth, Plymouth PL4 8AA, Devon, United Kingdom; \(^b\)Psychology Department, State University of New York at Binghamton, Binghamton, NY 13902-6000.

ikirsch@plymouth.ac.uk

Abstract: Although we are sympathetic to his central thesis about the illusion of will, having previously advanced a similar proposal, Wegner’s account of hypnosis is flawed. Hypnotic behavior derives from specific suggestions that are given, rather than from the induction, of trance, and it can be observed in 90% of the population. Thus, it is very pertinent to the illusion of will. However, Wegner exaggerates the loss of subjective will in hypnosis.

Hypnosis and will. In a manuscript that we submitted to Wegner in 1995, in his capacity of associate editor of Psychological Review, we also reached the conclusion that “volition is not an introspected content of consciousness, but rather an interpretation.” Our thesis was:
Commentary/Wegner. Précis of The illusion of conscious will

annexia, and hallucinations. Because these states are so striking, it was assumed for centuries that they must be due to a special or unusual condition. However, all of these altered states can be produced without the induction of hypnosis or any other special state. Instead of revealing the presence of a hypnotic trance, they disclose a normal human capacity to profoundly alter subjective experience. (Kirsch 2001, p. 795)

For this reason, there is no inconsistency between nonstate theories and data indicating brain changes accompanying the experience of hypnotic suggestions:

Finding physiological concomitants of this sort would be consistent with all theories, including socio cognitive theory. All subjective experiences are assumed to have physiological substrates (Hyland 1985). Thus, there is no reason why this should not be true of the subjective reactions to suggestions. (Kirsch & Lynn 1995, p. 855)

Ironic processes in hypnosis. We are sympathetic to Wegner’s analysis of ironic process and have extolled its clinical implications (Kirsch & Lynn 1999a). Nevertheless, a test of his application of ironic process theory to hypnosis has produced negative results (Kirsch et al. 1999). Based on the assumption that hypnotized subjects try to prevent responses from occurring as simple voluntary acts, Ansfield and Wegner (1996) proposed that while the intentional operating process is attempting to suppress the response, the ironic monitoring processes is searching for indications of it, thereby increasing the accessibility of suggested thoughts and movements. In this way, “the hypnotic state bypasses the ironies of mental control” (Wegner 2002, p. 311). If this were the case, cognitive load should enhance responsiveness to hypnotic suggestions. In fact, it does the opposite (Kirsch et al. 1999). Instead of enhancing responsiveness, cognitive load inhibits the ability to respond to suggestion, just as it does with nonhypnotic volitional behavior. Although inconsistent with the ironic process account of hypnotic behavior, this finding is consistent with the central thesis of Wegner’s book, the idea that the distinction between volitional and automatic behavior lies in the subjective judgment of the individual, rather than in fact.

NOTE

1. The mistaken idea that social cognitive theories of hypnosis are based on faking may be related to Sarbin’s (1980) use of social psychological role theory to explain hypnotic behavior. It is important to note, however, that Sarbin referred to “role-taking” rather than “role-playing” to describe the determinants of hypnotic behavior and experience. People engage in multiple social roles (e.g., researcher, writer, teacher, parent, and spouse), and their behavior is altered as a function of which role they are in. These role-induced alterations in behavior occur automatically (i.e., without volitional planning) and are accompanied by corresponding alterations in experience. Thus, the effect of taking on a social role is not an indication that the person is faking; Wegner has taken on the role of a writer and we are in the role of reviewers, but we are not faking and we presume that Wegner is not faking either.

Experimental psychology cannot solve the problem of conscious will (yet we must try)

Joachim I. Krueger

Department of Psychology, Brown University, Providence, RI 02912.

Joachim.Krueger@brown.edu

http://www.brown.edu/Departments/Psychology/faculty/krueger.html

Abstract: According to the view that humans are conscious automata, the experience of conscious will is illusory. Epistemic theories of causation, however, make room for causal will, planned behavior, and moral action.

Humans often experience a state of conscious will prior to their own actions. Yet (and by definition), they remain unaware of the nonconscious mental processes that precede both. The conjunction of precedence, consistency, and exclusivity gives rise to the strong and stubborn conviction that will can cause action. Wegner (2002) considers this illusionary; arguing that only the antecedent nonconscious processes are causal, whereas the belief in conscious will is epiphenomenal. The view that humans are conscious automata has a long history, as Wegner amply documents. He then reviews experimental findings that show how nonconscious events can predict actions, and how the belief in the causal power of conscious will can be strengthened or weakened. Is this evidence sufficient to validate the claim that conscious will is epiphenomenal?

Some theorists view causation in ontic terms, meaning that causal processes are properties of the world independent of the state of human knowledge (Salmon 1984). Other theorists view causation in epistemic terms, meaning that causation is a matter of epistemic inferences drawn from available data (Russell 1948). Wegner’s characterization of actions as having true (nonconscious) psychological causes suggests an ontic view, whereas his characterization of introspective perception of will suggest an epistemic view. As tools for making inductive inferences, psychological experiments operate within an epistemic framework, leaving ontic claims to those philosophers who wish to make them. Inferences about causation are just that: evidence-based speculations regarding how observed episodes (e.g., behavior) may be best explained. When experimenters generate such explanations, they do what humans always do: they use the cues of precedence, consistency, and exclusivity to strengthen or weaken certain ideas regarding what leads to what (Hume 1777/1900). In the epistemic view, nonconscious events are no more real than conscious events. Nonconscious events may take temporal precedence over conscious ones, and, by definition, only the experimenters know about them. This privileged access to earlier information tempts experimenters to dismiss subjective explanations on the grounds of the “We-know-more-than-you-do theory” (Krueger & Funder 2004). However, the experience of conscious will has the advantage of being closer to the time of action, and often the last event preceding an effect is seen as the most potent cause (Spelmann 1997). The experience of conscious will represents the aggregated activities of antecedent nonconscious activity. Just like conscious will, nonconscious mental activity at any given level of aggregation can be discounted as being the “true cause” of action because there is always another and more molecular level of activity preceding it.

Wegner claims that chains of nonconscious causes culminate in both the perception of conscious will and overt action. The implication that conscious will is a spurious cause of action can be viewed in light of two kinds of theory. Given regularity theories, nonconscious events (N) entail both the experience of conscious will (W) and overt action (A). W is judged epiphenomenal because it does not cause A. Yet the same view implies that W is necessary for A to occur, for if W were denied, so would N (modus tollens), and without N, no A (unless something else causes A). In this view, W is a necessary though non-causal antecedent of A. W might still be viewed as being epiphenomenal if it had no other effects. Wegner allows such effects, however, namely a sense of morality and responsibility. With the suggestion that without W, “memory for the emotional consequences of our actions would not guide us in making moral choices in the future” (Wegner 2002, p. 341), the epiphenomenality hypothesis collapses. W re-enters the causal chain, leading people to do the right thing some of the time.

Given probabilistic theories, N makes A more probable regardless of W. The path from N to A “screens off” any effect of W on A (Reichenbach 1956). Inductive experimental research thus needs to show that the path from W to A is spurious, but the idea of proving a null hypothesis remains controversial (Krueger 2001). Nevertheless, the research Wegner cites is dedicated to controlling various N and showing their effects on A. This work is convincing inasmuch as there cannot be parallel work in which W is an independent variable. To allow conscious will, experimenters would have to yield control of the independent variable, and these studies would no longer be experimental. Subjects cannot take control of this variable because they cannot separate their wish to
Commentary/Wegner: Précis of The illusion of conscious will

Free will for everyone — with flaws

George Mandler
Department of Psychology, University of California, San Diego, La Jolla, CA 92093-0109. gmandler@ucsd.edu

Abstract: Wegner's refutation of the notion of a conscious free will is addressed to a general reader. Despite a wide ranging and instructive survey and a conclusion acceptable to current psychological thinking, it is flawed by terminological confusions and lack of attention to relevant evidence and previous psychological approaches. It is suggested that psychology best addresses such factors as cognitive dissonance, attitude change, and locus of control as well as various variants of decisions such as "rational decision," "snap decision," "random choice," and "coerced choice." All of these are accompanied by "experienced will." Westcott offers a flow chart of the precursors of such experienced will that combines historical and current determinants and cognitive activity (including attention, valuation, and criterion setting). The final result is remarkably similar to Wegner's conclusions about empirical will.

I mention the paper that Kessen and I presented in 1974 primarily in order to make an additional argument. We noted that whereas free will is a human construction rather than a fact of existence, a belief in free will is still probably a desirable state of affairs. The belief that one is free to choose from among different alternatives generates a delay in thought and action that brings more alternatives to the fore, and strengths among them may change in the light of evidence. Such a delay "is likely, though not certain, to bring some increment to the quality of the final choice" (Mandler & Kessen 1974, p. 316). We also suggested that as young children discover that their actions influence their environment, they develop a theory of personal efficacy that contributes to the belief in voluntary control. Our suggestions add in small part to Wegner's notion in Chapter 9 that the experience of free will acts to organize our experience of our own agency.

Wegner's final chapter starts with a well-argued discussion of the relationship between conscious willing and determinism, and makes interesting contributions to the advantage of conscious will in providing a sense of authorship and of achievement. Finally, while Wegner's distinction between conscious and empirical will is useful, what is missing is a disciplined discussion of the empirical will. Wegner (as well as other writers such as Westcott) leaves us with a complex menu of possible contributores to intentional, directed action — but no roadmap, no recipes. Maybe it would be best to forget about the problem of will altogether. Now that we understand what the subjective feeling of willing is about, we can return to our major problem: to understand, explain, and predict human thought and action. Will, in general, is too easily confused with conscious, illusory will. It also has unfortunate links with theories of the will associated with national socialist Germany (Mandler 2002). I would prefer to define conscious will in terms of Wegner's explanation, and get on with the work of psychology without extraneous baggage, such as attempts to define a determinist will.
Commentary/Wegner: Précis of The illusion of conscious will

Inferences are just folk psychology

Thomas Metzinger
Philosophisches Seminar, Johannes Gutenberg-Universität Mainz, D-55099 Mainz, Germany. metzinger@uni-mainz.de http://www.philosophie.uni-mainz.de/metzinger/

Abstract: To speak of “inferences,” “interpretations,” and so forth is just folk psychology. It creates new homunculi, and it is also implausible from a purely phenomenological perspective. Phenomenal volition must be described in the conceptual framework of an empirically plausible theory of mental representation. It is a non sequitur to conclude from dissociability that the functional properties determining phenomenal volition never make a causal contribution.

I have offered an alternative interpretation of some of Dan Wegner’s most relevant data elsewhere (Metzinger 2003, p. 506ff), and will confine myself to three conceptual points here. Wegner’s project could be further strengthened by eliminating an omnipresent version of the mereological fallacy, by adopting an empirically plausible theory of mental representation, and by avoiding certain kinds of non sequiturs.

In his laudable attempt to describe and more carefully analyze the functional architecture of phenomenal volition Wegner frequently employs personal-level concepts and predicates like “interpreting” (e.g., thoughts as causes [Wegner 2002, p. 64ff]), “inference” (e.g., of an apparent causal path, p. 68ff), or “control” (e.g., mental control, p. 310ff). The author uses such predicates to predicate and concepts simultaneously on personal and subpersonal levels of description. At one time he speaks of the whole person as interpreting, for instance, her own thoughts as causes, and at another time of an “interpretive system” on the subpersonal level (e.g., as a course-sensing mechanism, p. 317); at one time he analyzes the person as a whole exerting mental control, at another he talks about a “controlling apparatus” (e.g., p. 312), and so forth. The subpersonal readings are all fallacious: Brains – or functional subsystems of brains – don’t interpret anything, they don’t make any inferences, and they don’t exert control. Only whole persons can be directed at the meaning of certain sentences (or of sentences describing chains of internal events), thereby attempting to interpret them. Only whole persons could establish inferences between mentally represented propositions. And, only whole persons can be directed at the fulfillment conditions defining certain goal-states, that is, only whole persons can truly make an attempt at controlling a certain state of affairs.

The deeper problem in the background is that one needs an empirically plausible and conceptually coherent theory of mental representation to successfully describe the architecture of phenomenal volition on a subpersonal level, that is, without committing the homunculus fallacy. Daniel Wegner does not develop such a theory, but assumes that apparent mental causation results from “interpretations” and “inferences.” Brains, however, are not inference machines, but associative engines (see, e.g., Clark 1989; 1998). Probably brains are even more than that, namely, complex dynamical systems exhibiting something like a “liquid” architecture. It has now become overwhelmingly plausible that such systems do not exhibit a critical property which Ramsey et al. (1991) have called “propositional modularity”: the fact that they represent propositional content in a way that makes individual units functionally discrete, semantically interpretable, and endowed with a distinct causal role. In this light the “inferences” underlying apparent mental causation are a leftover piece of folk psychology that has to be substituted by a suitable subsymbolic/dynamical story. Second, “inferences” and “interpretations” also are phenomenologically implausible, because none of us actually subjectively experience ourselves as drawing inferences and interpreting syntactical structures before having the conscious experience of will. They are leftover pieces of folk phenomenology. As a matter of fact these two points can now be seen as a new constraint for all candidate theories of mental representation: Are they able to accommodate a fine-grained and subsymbolic analysis for the architecture of conscious volition, functionally as well as phenomenologically?

Every form of phenomenal content has at least one minimally sufficient neural correlate (Chalmers 2000). This is true of every instance of consciously experienced volition too: For every such experience there will be a minimal set of neurofunctional properties that reliably activates it and which has no proper subset that would have the same effect. Many philosophers would even argue that every single instance of phenomenal volition is token-identical to this very correlate.

Interestingly, in a given system, every single overt action has at least one such minimally sufficient neural correlate too. For every such action there will be a minimal set of neurofunctional properties that reliably brings it about, and which has no proper subset that would have the same effect. Dan Wegner has made a major contribution in showing how many situations there are in which behavior and phenomenal will can be dissociated in various ways, and what the parameters guiding such dissociations are. Given his data, it is a rational and plausible conclusion to assume that both kinds of sets of neurofunctional properties only loosely overlap. At times they can be instantiated in isolation. What does not follow is the proposition that – especially in nonpathological standard configurations – the functional properties determining phenomenal volition never make a considerable contribution to action control. This is a non sequitur.

What we have to distinguish are cases where apparent mental causation is mere appearance, and cases where appearance and mentally represented knowledge possibly coexist. In philosopher’s jargon, we need a criterion that allows us to distinguish between these cases when conscious will is only phenomenal content, and cases where episodic, intentional content is co-instantiated in the very same event. Let me give an example: In standard configurations the functional properties determining the fact that the experience of conscious will occurs could at the same time could be a subset of exactly those functional properties that make the self-organizing dynamics of a certain, ongoing motor selection process globally available, thereby adding flexibility, context-sensitivity, and accessibility. Contrariwise, if the experience of conscious will occurs could at the same time be a subset of exactly those functional properties that make the self-organizing dynamics of a certain, ongoing motor selection process globally available, thereby adding flexibility, context-sensitivity, and accessibility. Contrariwise, if the feeling of will could then be not an illusion, but, rather, a nonconceptual form of self-knowledge – that is, the introspective knowledge that one right now is a system undergoing the internal transformation just described.

Differentiating dissociation and repression

John Morton
Institute of Cognitive Neuroscience, University College London, London WC1N 3AR, United Kingdom. j.morton@ucl.ac.uk

Abstract: Now that consciousness is thoroughly out of the way, we can focus more precisely on the kinds of things that can happen underneath. A contrast can be made between dissociation and repression. Dissociation is where a memory record or set of autobiographical memory records cannot be retrieved; repression is where there is retrieval of a record but, because of the current task specification, the contents of the record, though entering into current processing, are not allowed into consciousness. I look at hypnotic amnesia and dissociative identity disorder in relation to this contrast.

Wegner has set up a framework within which phenomena such as post-hypnotic amnesia and dissociative identity disorder (DID) sit very comfortably, even though the paradoxes are heightened. Consider, if conscious will is an illusion, then acting without the experience of consciously willing one’s actions can be seen as realism.

I will focus on amnesia in the contexts of hypnosis and DID. As a cognitive psychologist, I am interested in the nature of the amnesia. I regard the autobiographical memory system as separate

from a central cognitive processor and separate from the buffer store that services this processor. I also consider that the material in the buffer is not automatically accessible to consciousness. Such a position sits well with the framework that Wegner has so elegantly laid out.

There are two major ways in which a person may fail to consciously retrieve an autobiographical memory. The first kind of problem is that the memory record in question cannot be accessed. There are a number of ways in which this could happen (cf. Mortonet al. 1985), but the outcome is that the material in the record does not arrive in the buffer store. It cannot, then, in this model, influence behaviour. The second kind of problem is that the material arrives in the buffer store, will be subject to some processing, and can influence behaviour. However, some mechanism different from the material being made available to conscious processing.

How might one distinguish these two broad classes? Consider the following experiment: You perform a free association task with a participant. Then, after some manipulation or other, you perform the identical free association task, using exactly the same stimulus words in the same order. What will be the influences on the outcome the second time through? Roughly speaking, there will be two influences. The first will be the priming of the responses given the first time around due to activation remaining in the perceptuo-semantic-motor system. Let me call this implicit priming, for short. The second will be the memory record of the first run through the task. Now, suppose that the experimental manipulation involves a hypnotic suggestion that the first task be forgotten. In principle, the forgetting might be achieved in either of the two ways outlined above. If the record of the first task is inaccessible, then its only influence would be that of the implicit priming. There should be a lot of repeat responses and they should be faster than in the first run. In fact, when the study was run (Mortonet al. 2000) hypnotised subjects took longer on their second run than on their first run (N = 14; mean response time first run, 1.4 seconds; second run, 1.9 seconds). They also produced a mean of 7.9/15 different responses on the second run. These figures compare with a control group who were encouraged to give the same responses who were faster on the second run (1.18 sec vs. 1.1 sec; 1.7/15 repeat responses). The decrement in the hypnotised group could only have come about if they had retrieved the information concerning the first run and used it to monitor their responses during the second run. In other words, they had repressed the information.

Consider the same experiment run with participants diagnosed with DID. The alters of interest are those who claim, directly or indirectly, that they have no knowledge of the activities of the other alters. You run the free association experiment with one alter, switch to a second alter who claims to know nothing of what has just happened and run the experiment again. If the second alter has no access to the memory records laid down by the first alter — what I call dissociation — then (individual differences be- between the alters permitting) the second alter should be faster than the first. I have run this experiment on three DID patients to date. With DID1 the response times were 1.7 secs for alter 1 and 1.5 secs for alter 2 in one experiment and 1.4 to 1.2 secs in a second experiment. With DID2 the times were 3.1 to 2.7 seconds. These preliminary data indicate that these DID patients are dissociating — in contrast to exhibiting repression like the hypnotised subjects. Finally, DID3 reported an increase in response time from 1.7 to 1.9 seconds. This patient had failed to give clear evidence of dissociation in other tests, and so the association task seems to have some diagnostic utility.

Wegner (2002) supposes that DID switching is equivalent to rebooting a computer with a different operating system (p. 269). My own feeling is that DID switching is more like logging out and then logging in under a different user name, with a denial of access to the personal files of the other users, though with the same operating system and user programs. Otherwise we would not find priming.

Abstract: A causally efficacious conscious will is a small part of our everyday activities, but a part that deserves to be recognized, studied, and cherished, perhaps as a fundamental, emotion- and conation-related, right hemispheric neuronal process. Such brain functions might be less in doubt if we consider all the pieces of the larger pie, especially those where our passions and desires reside.

Wegner (2002) offers fascinating journeys through carnival aspects of the human cognitive apparatus. He gracefully coaxes us to abandon a cherished belief: that our actions evolve around willfulness within the conscious universe of the brain/mind. As long recognized, “the whole subject of unconscious cerebration . . . is pregnant with interest” even as some “draw what must be regarded as untenable and artificial distinctions between reality and resemblance in conscious and unconscious mental action. They suggest, if they do not assert, that purposive actions may possess a false appearance of ideation, a deceptive volition” (Lindsay 1879, p. 7). Wegner proceeds steadily in that direction, with modest conviction.

Should mind scientists finally agree that human thoughts cannot voluntarily intend actions? Not at all, if only a modest slice of pie is presented as the whole, especially since our left hemisphere “interpreter” is so commonly a “confabulator” (Turnbull & Solms 2004).

Many credible scenarios are left. What about the measured actions and potentially “ironic” willfulness of our self-absorbed and pessimistic right-hemispheric “realist” (Davidson & Hugspeth 1995)? What about the consciousness of those ancient emotional operating systems that generate our animalian intentions-in-action (Panksepp 2003a; 2003b)? Although the extroverted left hemisphere enjoys a good story and pontificates obsessively to grease the social wheels, might other brain areas be more adept at provoking self-consciously motivated actions?

Free will may be more critically linked to imagery-attuned functions of the right hemisphere, in close touch with periconscious subcortical emotional functions. Wegner does peer behind William James’s “illusion of screens” as he exhibits menageries of peculiar mental proclivities. However, he avoids our deeper animalian nature, wherein persistent desires and willfulness are not just social constructions, but animalian attributes of our dopamine fired seeking urges (Panksepp 1998a). The feeling of conation, resurrected briefly as a willful “cognitive emotion,” is a promising candidate from that periconscious realm. Such intention-generating processes are not deeply unconscious, although they often fade under glaring Hollywood-like screens of perceptual and linguistic consciousness.

Here is my recent encounter with the pure feeling of free will: During surgery under spinal blockade, I could no longer feel myself voluntarily wiggle my feet, but move they did, predictably, verifiable by looking. That spooky feeling of effference, without somatosensory/proprioceptive feedback, was part of my volitional apparatus. Might the periconscious conative borderland between our animalistic “intentions in action” (affective consciousness) and our human “intentions to act” (cognitive consciousness) be where...
Commentary/Wegner: Précis of The illusion of conscious will

the neural roots of our admittedly beleaguered free-will are sought?

With massive hierarchical layering of affective, conative, and cognitive controls, each with some consciousness (Panksepp 2003a,b), I can imagine how “free will” emerged in our thoroughly neuronal “mind machines.” Creatures that can conceptualize alternative paths for fulfilling desires, have brain processes that can intend to act much more than organisms thoroughly captivated by their affectively rich intentions in action. With a robust emotional action apparatus devoted to instinctually actuating our many needs/desires and emotional selves, and several layers of thoughtful decision-making (perceptual and linguistic), we have a solid grounding for willfulness as well as for dissociative personality disorders. In contemplating such deep evolutionary issues, I question classical interpretations of “Libet’s paradox,” too readily accepted by Wegner.

Libet’s (1999) paradox looses its impact if central psychomotor commands help the sensory-perceptual apparatus to harvest external action-related information. Brain action coordinates are foundational for perceptual maps (Sparks 1988) and the core emotional self (Panksepp 1998a; 1998b). If evolution prioritized action over perception in the hierarchical layering of behavioral control processes within the brain, perceptual correlates (e.g., Libet’s timing of willed actions) should occur after one has internally initiated willed actions (Panksepp 2003b). From this vantage, Libet’s data is consistent with the causal efficacy of intentionality. If we forget that action dynamics can have feelings of their own (Panksepp & Gordon 2003), our linguistic consciousness can be captivated, too, easily, by left-hemispheric semantic paradoxes. Might Wegner’s “false dichotomy” between free will (purportedly a “feeling”) and determinism (surely a “process”), be blended toward flexible, dual-aspect monism by recognizing that basic feelings are processes of the brain (Panksepp 1998a)?

Might emotional-conative instinctual consciousness provide self-representational infrastructures for the higher cognitive apparatus? If we consider the general purpose appetitive motivations feeling system that actuates our many specific needs and desires, and recognize how this “primal will” is linked to a complex cognitive-learning apparatus that can generate habitual and delusional ways of acting in the world (Izumoto & Panksepp 1999; Panksepp 1998a), we can appreciate, along with Wegner, how little free will we commonly exhibit. However primitive the intentions in action are that control much of our behavior (emotional instincts, seemingly with a mind of their own), they do intervene at higher reaches of the brain that provide recursive controls that amount to intentions to act in ways not ordained by our insistent basic needs and emotions. To understand free will, we must fathom our higher affective-conative apparatus, perhaps more right than left hemispheric.

Free will is surely dependent on a deep sense of self-awareness, which arises from emotion-cognition interaction zones like the insular, cingulate, and medial frontal cortices, far from language areas (see e.g., Kanpe et al. 2003; Kelley et al. 2002; Kircher et al. 2000; Wicker et al. 2003). While many other animals have simpler forms of consciousness, we humans have levels of self-awareness that allow us to voluntarily facilitate certain actions and also to inhibit subcortical emotional urges (Liotti & Panksepp 2004). Yes, our motor apparatus readily becomes habitual, like a well-oiled cruising machine that only needs occasional steering by higher intentions, but that useful automaticity does not contradict the existence of voluntary willfulness.

If self-referential awareness and conation allows us some free will (and Wegner is surely right that we do not exercise it as much as some imagine), then future progress on the topic will require more sophisticated neuropsychological research. Semantic and behavior-only analyses cannot resolve these issues adequately. Even though minds are indeed nothing more than incredibly complex neuronal/glial/body machines, evolutionarily designed to operate in complex environments, the emergent feeling process of conscious will, hierarchically reconceptualized, can subsist within the complex, multi-tiered cognitive structures that are grounded on our emotionally rich animalian motivations and desires. Such abilities set us apart from most other animals. Of course, premature closure on such topics of ultimate concern would be foolish, even if for no other reason than that future generations also need to endlessly debate these scientifically unfathomable and perhaps unfathomable neuropsychological issues. I thank the author for a stimulating and provocative read.

The illusion of explanation: The experience of volition, mental effort, and mental imagery

Zenon Pylyshyn
Center for Cognitive Science, Rutgers University, New Brunswick, Piscataway, NJ 08854-8020. zenon@ruccs.rutgers.edu http://ruccs.rutgers.edu/faculty/pylyshyn.htm

Abstract: This commentary argues that the “illusion” to which Wegner refers in The Illusion of Conscious Will is actually the illusion that our conscious experience of mentally causing certain behaviors explains the behavior in question. It is not the subjective experience itself that is illusory, but the implied causal explanation. The experience of “mental effort” is cited as another example of this sort of illusion. Another significant example is the experience that properties of the representation of our mental images are responsible for certain patterns of behavior observed in mental imagery experiments. Examples include the increase in reaction time found when details are reported from smaller images or when attention is switched between different places and features (imagined as further apart than they are) within a single image. These examples illustrate the nature of the “illusion” involved: It is the illusion that certain observed regularities occur because of the content of the experience, as opposed to the converse – that experience has the content it does because of what the person figures out would happen in the imagined situation.

Wegner (2002) presents an excellent case for the view that when we experience ourselves as deciding to act or as intending some action, what we experience is not the actual cause of the subsequent behavior. This is not quite the same as claiming that the conscious experience of will is illusory, as suggested by Wegner’s title. It’s not that we are deceived about how things seem to us; what is illusory is that how things seem to us often feels like an explanation of the causes of the behavior. Many of our conscious experiences are experiences of causing some pattern of behavior and it is this attribution that is illusory, not the experience itself. Perhaps one should say that what is illusory is the way that conscious contents often appear to explain one’s actions (see Pessoa et al. 1998, for more on what they call the “analytical isomorphism” assumption).

It should come as no surprise that we rarely experience the causes of our behavior. People have no more conscious access to the information processing that underlies their behavior than they do to the biochemical or neural processes that instantiate them, nor do they typically even have conscious access to the tacit knowledge, implicit perceptions, and inferences involved in their own cognition. A case similar to the one Wegner makes for the experience of will can be made for many other conscious experiences, such as the experience of mental effort, mentioned briefly in the book. What we consciously experience as “more effort” could not correspond to something like using more of an information-processing resource (e.g., more operations, more storage capacity, etc.), even if it might sometimes be correlated with such quantities. One reason is that the experience of effort is affected by our beliefs – including how hard we believe a problem to be, how much we dread it, and how anxious or worried we are about being able to carry it out (witness the case of “math anxiety” which results in the experience of mathematical problems requiring a great deal of “mental effort”).

Given the problematic role of conscious contents in cognition, one might wonder what exactly our conscious experiences do reveal. It would be odd if our experience had nothing to do with the
processes that cause the behaviors. At least one line of research (that on the interpretation of experimental studies of mental imagery, e.g., Pylyshyn 1981) has provided evidence that the content and dynamics of people's consciously experienced imagery is a result of, rather than a cause of, what people know about the situation being imagined (as well as the way that the task of imagining is interpreted).

Studies of mental imagery provide a clear example of the illusory nature of explanations based on the conscious experience of one's cognitive process (for a detailed discussion of this issue, see Pylyshyn 2003b). One's experience of the form of images and of why they unfold the way they do provides one of the most misleading sources of explanations of mental processes. The experience of "seeing" events unfold in one's "mind's eye," and thereby of seeing why some operations are more difficult or take longer than others, is so compelling that it is almost impossible to discount. The fact that the conscious experience of visual imagery is similar to that of perception (presumably because of the involvement of some of the same brain mechanisms) suggests to many people that the representations involved in imagery must themselves resemble the content of the experience (viz., that they consist of picture-like displays) or that images are constrained in their dynamics by principles similar to those that govern the world being imagined (e.g., that they "rotate" while rigidly retaining their shape). Yet the inference from the form of the experience of imagining to the picture-theory (or a theory that claims we have a dynamic model of the world in our head) is based on unsupported assumptions, such as that a brain state responsible for the conscious experience of seeing must itself resemble what is seen, or that the brain is so constituted that images are required to follow principles similar to those that govern processes in the physical world.

As in the case of experienced volition, our conscious experience appears to provide a natural explanation for why certain behaviors occur. In mental imagery we not only have the experience of "seeing" but we also have the experience that certain patterns of the ensuing behavior are caused by properties of the representations that we consciously experience. For example, it seems clear why it takes us longer to report details in a "small" image than in a "large" one; our experience tells us that this is because the details are "harder to see" when the image is smaller. Similarly, it seems obvious why it takes us longer to switch our attention between two imagined objects when they are imagined as being further apart; our conscious experience shows that this is because attention takes longer to travel a greater distance across the surface of the image. Likewise, it is no puzzle why we find it more difficult to see the outer edges of our mental image; our experience shows us that this is because our "mind's eye" has a certain visual angle and when things get near the periphery they are harder to discriminate, just as things are harder to see in the periphery of vision. It thus seems that many properties of mental imagery, including why certain results are obtained in imagery experiments, can be explained by simply attending to the experience and seeing for yourself how the process happens.

However, the explanations suggested by conscious experience can easily be shown to be specious in examples such as the ones cited above. Even though our causal mental process may go through a sequence that corresponds to the sequence that we experience, it does not in general proceed that way for the reasons suggested by the conscious experience (for more on this, see Pylyshyn 2002, 2003a). The way our imagery unfolds—the sequence it goes through when we imagine certain events—is consistent with objective measures such as reaction times, but it cannot explain them. The experience of taking longer to scan greater imagined distances does not explain the reaction time observations, because the principle that it takes longer to travel a greater distance applies only to real motion over real physical distances, not to phenomenological motion, which can follow any principle one wishes (try imagining that your attention hops from place to place in your image without taking time that increases with distance). The real reason that our imagery goes through the sequence it does is, in many cases, simply that we make it go through that sequence because that is the sequence we expect in the situation we are imagining. To imagine something means to recreate what one believes would happen in the situation one is imagining. In other words, what we experience as arising from properties of the image itself is actually a consequence of our knowledge of how things would work in the imagined world. Evidence for this is that if we change what people believe would happen in the imagined situation, the observations also change predictably (see Pylyshyn 1981 for examples). What explains the behavior in these cases is not some principle that governs the dynamics of our image, as suggested by the conscious experience of watching the imagery unfold autonomously in one's mind's eye, but rather our (generally tacit) knowledge of the situation we are imagining (together with our psychophysical ability to simulate the sequence). It is in this sense that the conscious experience of mental imagery might be viewed as "illusory," though a better way to characterize it is that the experience of mental imagery provides a misleading explanation of why certain patterns of behavior occur.

A social psychologist illuminates cognition

Amir Raz and Kim L. Norman

*Columbia University College of Physicians and Surgeons and New York State Psychiatric Institute, New York, NY 10032; 2) Barnard College, Department of Psychology, New York, NY 10027; ar2241@columbia.edu

Abstract: Sprinkled with humor, social psychology illuminates cognition in Wegner's beautifully written and cleverly crafted book. However, scantly exploiting such themes as psychopathology, development, and neural correlates of consciousness, Wegner's account does not fully project into cognitive neuroscience. Broaching the topic of self-regulation, we outline neurocognitive data supplementing the notion that voluntariness is perhaps more post-hoc ascriptions than bona fide introspection.

Combining phenomenology with empirical data, Wegner, a social psychologist, skillfully elucidates the relationship between willed action and its underlying representations, taking the reader from the labyrinths of parlor magic into the realms of hypnosis. It is possible to gain insights into both healthy and pathological function by examining the healthy individual under atypical conditions. So powerful is the greater distance across the surface of the image. Likewise, it is no puzzle why we find it more difficult to see the outer edges of our mental image; our experience shows us that this is because our "mind's eye" has a certain visual angle and when things get near the periphery they are harder to discriminate, just as things are harder to see in the periphery of vision. It thus seems that many properties of mental imagery, including why certain results are obtained in imagery experiments, can be explained by simply attending to the experience and seeing for yourself how the process happens.

However, the explanations suggested by conscious experience can easily be shown to be specious in examples such as the ones cited above. Even though our causal mental process may go through a sequence that corresponds to the sequence that we experience, it does not in general proceed that way for the reasons suggested by the conscious experience (for more on this, see Pylyshyn 2002, 2003a). The way our imagery unfolds—the sequence it goes through when we imagine certain events—is consistent with objective measures such as reaction times, but it cannot explain them. The experience of taking longer to scan greater imagined distances does not explain the reaction time observations, because the principle that it takes longer to travel a greater distance applies only to real motion over real physical distances, not to phenomenological motion, which can follow any principle one wishes (try imagining that your attention hops from place to place in your image without taking time that increases with distance). The real reason that our imagery goes through the sequence it does is, in many cases, simply that we make it go through that sequence because that is the sequence we expect in the situation we are imagining. To imagine something means to recreate what one believes would happen in the situation one is imagining. In other words, what we experience as arising from properties of the image itself is actually a consequence of our knowledge of how things would work in the imagined world. Evidence for this is that if we change what people believe would happen in the imagined situation, the observations also change predictably (see Pylyshyn 1981 for examples). What explains the behavior in these cases is not some principle that governs the dynamics of our image, as suggested by the conscious experience of watching the imagery unfold autonomously in one's mind's eye, but rather our (generally tacit) knowledge of the situation we are imagining (together with our psychophysical ability to simulate the sequence). It is in this sense that the conscious experience of mental imagery might be viewed as "illusory," though a better way to characterize it is that the experience of mental imagery provides a misleading explanation of why certain patterns of behavior occur.

A social psychologist illuminates cognition

Amir Raz and Kim L. Norman

*Columbia University College of Physicians and Surgeons and New York State Psychiatric Institute, New York, NY 10032; 2) Barnard College, Department of Psychology, New York, NY 10027; ar2241@columbia.edu

Abstract: Sprinkled with humor, social psychology illuminates cognition in Wegner's beautifully written and cleverly crafted book. However, scantly exploiting such themes as psychopathology, development, and neural correlates of consciousness, Wegner's account does not fully project into cognitive neuroscience. Broaching the topic of self-regulation, we outline neurocognitive data supplementing the notion that voluntariness is perhaps more post-hoc ascriptions than bona fide introspection.

Combining phenomenology with empirical data, Wegner, a social psychologist, skillfully elucidates the relationship between willed action and its underlying representations, taking the reader from the labyrinths of parlor magic into the realms of hypnosis. It is possible to gain insights into both healthy and pathological function by examining the healthy individual under atypical conditions. So powerful is the greater distance across the surface of the image. Likewise, it is no puzzle why we find it more difficult to see the outer edges of our mental image; our experience shows us that this is because our "mind's eye" has a certain visual angle and when things get near the periphery they are harder to discriminate, just as things are harder to see in the periphery of vision. It thus seems that many properties of mental imagery, including why certain results are obtained in imagery experiments, can be explained by simply attending to the experience and seeing for yourself how the process happens.

However, the explanations suggested by conscious experience can easily be shown to be specious in examples such as the ones cited above. Even though our causal mental process may go through a sequence that corresponds to the sequence that we experience, it does not in general proceed that way for the reasons suggested by the conscious experience (for more on this, see Pylyshyn 2002, 2003a). The way our imagery unfolds—the sequence it goes through when we imagine certain events—is consistent with objective measures such as reaction times, but it cannot explain them. The experience of taking longer to scan greater imagined distances does not explain the reaction time observations, because the principle that it takes longer to travel a greater distance applies only to real motion over real physical distances, not to phenomenological motion, which can follow any principle one wishes (try imagining that your attention hops from place to place in your image without taking time that increases with distance). The real reason that our imagery goes through the sequence it does is, in many cases, simply that we make it go through that sequence because that is the sequence we expect in the situation we are imagining. To imagine something means to recreate what one believes would happen in the situation one is imagining. In other words, what we experience as arising from properties of the image itself is actually a consequence of our knowledge of how things would work in the imagined world. Evidence for this is that if we change what people believe would happen in the imagined situation, the observations also change predictably (see Pylyshyn 1981 for examples). What explains the behavior in these cases is not some principle that governs the dynamics of our image, as suggested by the conscious experience of watching the imagery unfold autonomously in one's mind's eye, but rather our (generally tacit) knowledge of the situation we are imagining (together with our psychophysical ability to simulate the sequence). It is in this sense that the conscious experience of mental imagery might be viewed as "illusory," though a better way to characterize it is that the experience of mental imagery provides a misleading explanation of why certain patterns of behavior occur.
Commentary/Wegner: Précis of The illusion of conscious will

of the volitional status of suggested behavior have not only become the subject of passionate debate but constitute the crux differentiating theories of hypnosis (Kirsch & Lynn 1998a). Hypnosis data drawing on cognitive science and neuroimaging have provided significant insights into this conundrum.

It is not a coincidence that some practitioners prefer the term “self-hypnosis” to “hypnosis” (cf. Olness & Kohen 1996). Participants in hypnosis studies generally wish to be hypnotized and therefore consent to fill the hypnotic role and follow suggestions. Their compliance differs from that of a voluntary response to a request in that they must make plans not only to execute a suggested movement, but also to concurrently interpret the movement as non-volitional. Indeed, there are data supporting this mental process (Silva & Kirsch 1992). However, whether or not hypnotic responses are intentional, it is important to remember that they are experienced as involuntary by the subject. As it is likely that these responses are a product of both intentional and automatic elements, the issue becomes more a question of whether the response is elicited intentionally or automatically (e.g., Raz & Shapiro 2002).

There are data showing that highly hypnotizable individuals can eliminate involuntary and ballistic effects (e.g., Stroop interference) following a specific posthypnotic suggestion (MacLeod & Sheehan 2003; Raz et al. 2002; 2003b; Schatzman 1980). When they do, specific brain changes related to this effect occur (Raz, in press). Furthermore, there are now genetic findings concerning individual differences that might relate to the distinction between highly and less hypnotizable people (Raz et al. 2003a; in press-a; in press-b) as well as evidence that hypnotic inductions might lead to “behavioral lesions” reminiscent of actions following verbal lesions (e.g., stroke) (Raz, in press). Indeed, the heritability of hypnotizability is among the highest of any psychological individual-difference measure identified to date (Morgan 1973; Morgan et al. 1970) and neuroimaging findings associated with such hypnotic and attentional modulations consistently implicate differential activation patterns in the anterior cingulate cortex (ACC) (Fan et al. 2003; Raz et al. 2003b; in press-b; in press-a).

A popular theory of cognitive control proposes that the ACC is part of a network involved in handling conflict between neural areas (see Fig. 1). While some researchers view the ACC through the lens of a conflict-monitoring model (Botvinick et al. 2001; Cohen et al., 2000), others construe it as a regulation model engulfing broader processes of consciousness and self-regulation, including executive attention and mentation (Bush et al. 2000).Consistent with the importance of the ACC to normal conflict monitoring, there are syndromes of abnormal agency that occur with extensive lesions of the ACC and associated midline frontal cortex whereby a patient interprets the actions as caused by an outside force (Goldberg 1985). The ACC is well-equipped to mediate between limbic motivational influences and the adjacent supplementary motor areas, and lesions associated with ACC and medial frontal regions have been documented to produce akinetic syndromes, in which patients do not engage in actions despite being quite capable of doing so (Damasio & Van Hoesen 1983). With their ACC impaired, these patients appear to lack motivation to act. Towards this end, psychosurgery sometimes aims for the ACC to alleviate chronic pain or decrease the symptoms of anxiety, as such interventions typically decrease the patient’s concern over life problems (Rainville et al. 1997). The illusion of conscious will can be also harnessed towards a low-cost and noninvasive therapeutic means. For example, hypnotic interventions have been used to alleviate tic symptoms in individuals diagnosed with Tourette syndrome (TS) (Crawford 1992; Culbertson 1989; Cohen 1995; Cohen & Botts 1987; Lindner & Stevens 1967; Young & Montano 1988; Zahn 1987). Hypnotic suggestion is believed to engage self-regulatory mechanisms (Ray & Tucker 2003), and, whereas effortful control can evanescently suppress TS symptomatology, rendering self-regulation a lens by which to view TS formulation, the fact that volitional as well as involuntary control of behavior can be interrupted and modified by external suggestion proposes that, at least under appropriate conditions, hypnotic influence may engage mechanisms of control at an elementary level. By understanding the substrates of these processes, therefore, we may better understand not only the interesting phenomenon of conscious will, but mechanisms of self-regulation. This is particularly appealing in the context of human development, wherein studies have shown that the sense of control over actions becomes stronger with age. In this regard, studies of hypnotic susceptibility have repeatedly shown that children are more hypnotizable than adults (London 1965; Olness & Kohen 1996) and more readily attribute the cause of their actions to an external source, suggesting that the separation of action from authorship is perhaps more potent in younger age. The maturation of self-regulatory mechanisms across development is instructive in this sense, because prefrontal brain development reflects changes in perception of control over actions as well as thought and emotion and may lead to a more complete understanding of the correlates of conscious will (Bronsom 2000).

In conclusion, Wegner’s book is a delightful composition and a fine demonstration of how cognitive science can learn from the insights of an accomplished social psychologist. Although we would have liked to see a more rigorous treatment of relevant psychopathology and, particularly, data concerning the neural correlates of consciousness, books take time to prepare and some of the data we cite here were probably unavailable as Wegner was putting pen to paper. Apropos, Christof Koch’s latest, Quest for Consciousness (2004) nicely complements Wegner’s efforts on these points.
Conscious will in the absence of ghosts, hypnotists, and other people

Johannes Schultz*, Natalie Sebanz*, and Chris Frith†

*Wellcome Department of Imaging Neuroscience, University College London, London, WC1N 3BG, United Kingdom; †Max-Planck Institute for Psychological Research, 80799 Munich, Germany.

Abstract: We suggest that certain experiences reported by patients with schizophrenia show that priority, consistency, and exclusivity are not sufficient for the experience of willing an action. Furthermore, we argue that even if priority, consistency, and exclusivity cause the experience of being in control of one’s actions, there is no evidence that the author of an action, this does not mean that conscious will is an illusion.

Wegner (2002) discusses an impressive variety of phenomena demonstrating that when the three conditions priority, consistency, and exclusivity are met, an action feels willed, whereas when one or more do not apply, the cause of an action is attributed to forces other than the self. He convincingly shows that the feeling of conscious will can be erroneous, such that a person can either believe he was the author of an action even though he was not, or that he can believe he was not the author while in actual fact he was. The strongest version of Wegner’s claim would be that priority, consistency, and exclusivity are both necessary and sufficient for the experience of willing an action. However, we suggest that certain experiences reported by patients with schizophrenia show that priority, consistency, and exclusivity are not sufficient for the experience of willing an action.

Patients with delusions of control report that their actions, even quite trivial actions, are being controlled, not by themselves, but by some alien force. Patients report such abnormal experiences even though they have the prior intention to make the action, the action is made consistent with their intention, and there is no evidence of ambiguity about who is making the action. We have suggested elsewhere (Holroy & Frith 2004) that what is missing is an aspect of the feeling of what it is like to be in control of one’s actions; knowing what is going to happen and, at the same time, minimal awareness of the sensory consequences. Thus, will has a specific phenomenology in addition to the knowledge of authorship. We also argue that, even if priority, consistency, and exclusivity are sufficient for the experience of being the author of an action, this does not mean that conscious will is an illusion. The situations Wegner draws upon to claim that conscious will is simply an emotion of authorship are all very specific and differ in important ways from everyday settings. First, they are characterized by a lack of exclusivity, such that the intention to perform an action can either be attributed to oneself or another entity; be it a hypnotist, a ghost, or simply another person. Faced with a lack of exclusivity, we are likely to attribute authorship of an action to somebody else—unless priority and consistency are reinforced as in the “I Spy” study, wherein people are tricked into attributing to themselves an intention they never had. In everyday life, most of our actions and intentions can usually unambiguously be attributed to ourselves. Second, Wegner focuses on situations where intentions in action rather than prior intentions (Searle 1983) are at stake. He investigates the feeling of authorship in situations where one did not have a strong prior intention to perform a specific action. However, in everyday life, many of our actions seem to be the consequence of prior intentions that have been formed following conscious deliberation. A recent experiment (Lackner et al., in preparation) suggests that when a prior intention for an action has been formed, performance of the action is less susceptible to the influence of a distracter (a voice referring either to the action to be performed or an action not to be performed) than when the action is only accompanied by an intention in action. It seems that Wegner, in his remarkable study of the phenomenal will, has extended his conclusions slightly too far to include all kinds of intentions, and while his thought-provoking ideas explain cases of intentions in action, they do not explain prior intentions very well.

Finally, we suggest that from the finding that the phenomenal will can be illusory it does not follow that the empirical will, defined as “the causality of the person’s conscious thoughts as established by a scientific analysis of their covariation with the person’s behavior” (Wegner 2002, p. 14) is also an illusion. Although Wegner claims to address only the phenomenal will, he uses demonstrations of how the feeling of conscious will can be erroneous at times to draw conclusions about the empirical will, suggesting that all or most of our voluntary actions are caused by unconscious forces rather than conscious intentions. From the observation that the feeling of conscious will and actions are not causally related in certain specific conditions such as hypnosis, automatisms, and particular experimental settings, it does not automatically follow that conscious thoughts are generally not causally related to actions.

ACKNOWLEDGMENTS

Johannes Schultz and Chris Frith are supported by the Wellcome Trust. Natalie Sebanz is supported by the Max-Planck Gesellschaft.

Is the illusion of conscious will an illusion?

Robert J. Sternberg
PACE Center, Department of Psychology, Yale University, New Haven, CT 06520-8358. Robert.Sternberg@yale.edu www.yale.edu/rjsternberg

Abstract: This book is a tour de force in showing that what we believe to be actions dictated by conscious will are not, in fact, wholly dictated by conscious will. However, Wegner has fallen into the trap of making claims that go beyond his data to make his case more compelling and newsworthy. Psychology needs to be informed by common sense.

The Illusion of Conscious Will (Wegner 2002) is a wonderful book that shows that much of what we believe to be consciously-driven action is, in fact, more complexly driven than we are likely to think possible. For those who maintain an illusion of tight control, the book will be an eye-opener. For some notable examples, the foibles of Strom Thurmond, Nest Gingrich, Bill Clinton, Richard Nixon, and other very intelligent individuals may make clear, at least to some, that the strongest version of Wegner’s claim would be that priority, consistency, and exclusivity are both necessary and sufficient for the experience of willing an action. However, we suggest that certain experiences reported by patients with schizophrenia show that priority, consistency, and exclusivity are not sufficient for the experience of willing an action. Patients with delusions of control report that their actions, even quite trivial actions, are being controlled, not by themselves, but by some alien force. Patients report such abnormal experiences even though they have the prior intention to make the action, the action is made consistent with their intention, and there is no evidence of ambiguity about who is making the action. We have suggested elsewhere (Holroy & Frith 2004) that what is missing is an aspect of the feeling of what it is like to be in control of one’s actions; knowing what is going to happen and, at the same time, minimal awareness of the sensory consequences. Thus, will has a specific phenomenology in addition to the knowledge of authorship. We also argue that, even if priority, consistency, and exclusivity are sufficient for the experience of being the author of an action, this does not mean that conscious will is an illusion. The situations Wegner draws upon to claim that conscious will is simply an emotion of authorship are all very specific and differ in important ways from everyday settings. First, they are characterized by a lack of exclusivity, such that the intention to perform an action can either be attributed to oneself or another entity; be it a hypnotist, a ghost, or simply another person. Faced with a lack of exclusivity, we are likely to attribute authorship of an action to somebody else—unless priority and consistency are reinforced as in the “I Spy” study, wherein people are tricked into attributing to themselves an intention they never had. In everyday life, most of our actions and intentions can usually unambiguously be attributed to ourselves. Second, Wegner focuses on situations where intentions in action rather than prior intentions (Searle 1983) are at stake. He investigates the feeling of authorship in situations where one did not have a strong prior intention to perform a specific action. However, in everyday life, many of our actions seem to be the consequence of prior intentions that have been formed following conscious deliberation. A recent experiment (Lackner et al., in preparation) suggests that when a prior intention for an action has been formed, performance of the action is less susceptible to the influence of a distracter (a voice referring either to the action to be performed or an action not to be performed) than when the action is only accompanied by an intention in action. It seems that Wegner, in his remarkable study of the phenomenal will, has extended his conclusions slightly too far to include all kinds of intentions, and while his thought-provoking ideas explain cases of intentions in action, they do not explain prior intentions very well.

Finally, we suggest that from the finding that the phenomenal will can be illusory it does not follow that the empirical will, defined as “the causality of the person’s conscious thoughts as established by a scientific analysis of their covariation with the person’s behavior” (Wegner 2002, p. 14) is also an illusion. Although Wegner claims to address only the phenomenal will, he uses demonstrations of how the feeling of conscious will can be erroneous at times to draw conclusions about the empirical will, suggesting that all or most of our voluntary actions are caused by unconscious forces rather than conscious intentions. From the observation that the feeling of conscious will and actions are not causally related in certain specific conditions such as hypnosis, automatisms, and particular experimental settings, it does not automatically follow that conscious thoughts are generally not causally related to actions.

The title of the book implies that conscious will is a myth. Indeed, Wegner ends the book by stating that “the feeling of doing is how it seems, not what it is— but that is as it should be. All is well because the illusion makes us human” (p. 342). But is it an illusion?

I would argue that nothing in the book quite shows conscious will to be an illusion. Rather, it is part of a complex chain of events in which the conscious will does not necessarily come at the beginning of the chain. However, as Aristotle and everyone since who has studied causality has appreciated, causality always represents a complex chain of events. One can almost always ask for a cause one step further back in a causal chain. For example, why do people procreate? Because they want to? Because of their motivations? Because of their emotions? Because evolution drives them to? Because they want to? Because of their motivations? Because they want to? Because of their emotions? Because evolution drives them to? Because God willed them to? Because they are victims of their genes? The causal chain is long, and it is complex rather than linear. The fact that there may always be one step further back does not mean that causal value cannot be assigned to each step along the way. To argue otherwise is the ultimate in reductionism.

An example can be viewed in the case of the murders committed by Lee Boyd Malvo in and around the Washington, DC, area in 2002. It is uncontroversial that Malvo committed them. But why? Because he was under the dominating influence of John
Commentary/Wegner: Précis of The illusion of conscious will

Muhammad? Because he was psychopathic? Because he was a natural-born killer? The causal chain, as in most events, is long and complex. Unquestionably, research paradigms such as those used by Wegner would show that his conscious willing of the killings was not at the beginning of the causal chain. Was the jury therefore wrong in convicting him of murder and sentencing him to life in prison? The causal chain is complex. But one would shudder to think of Wegner testing at the trial that Malvo's will was only at some intermediate step in the causal chain, and that therefore, Malvo, as well as Muhammad, must be set free. They aren’t responsible for their actions because their conscious volition was at some midway stage of the decision process. Does Wegner or anyone else want to move to this position—that no one is responsible for his or her own actions? Do we want, in deciding what is, to spend our time deciding exactly what “is” means, as some powerful defendants would have us do?

I believe there is a general lesson here. Mischel (1968) once argued that research did not support the notion of personality traits. Jensen (1998) has argued that when all is said and done, general ability (g) pretty much captures all that is worth capturing in the study of intelligence. These claims seem, through common sense, off-base. Mischel (Mischel & Peake 1983) later backed off from his earlier claim. Perhaps someday Jensen or his disciples will back off from theirs. When the evidence of everyday experience suggests that the story told by psychological research is not quite right, we need to listen to it and consider the possibility that our paradigms are leading us astray, at least in our interpretation of conclusions. Wegner's research does not show conscious will to be an illusion. It shows it to be complexly determined. But I would suspect, or at least hope, that Wegner would not entirely exculpate Malvo or Muhammad on the argument that what they did was not the product of conscious will. Rather, the process was complex, but in the end, we must take responsibility for our own actions, however complexly determined they may be. The process by which Malvo committed the murders will not have started with conscious will. But conscious will could have kept Malvo from committing the murders. It didn’t. But he is culpable. And his culpability is no illusion, and it in no way makes him “human.”

Psychology, and science in general, have long been plagued by their failure to recognize fully the relevance of the Hegelian dialectic. Extreme theses and antitheses garner more attention, increase citation rates, and sell books. But they are rarely correct. In the end, questions that are originally formulated in terms of (false) dichotomies (e.g., “Is conscious will real or an illusion?”) usually end up being formulated in more complex terms that recognize some kind of synthesis in which a statement is not true or false, but rather true to some extent, under certain circumstances (Sternberg 1999).

None of my argument takes away from the value of Wegner's most impressive research project. But I believe it does call into question what may be an overly simplistic interpretation of the results. Conscious will is not an illusion. It is not a simple reality. It is part of a complex and multifaceted causal chain that cannot, in the end, relieve us of responsibility for our own actions.

Wegner’s “illusion” anticipated: Jonathan Edwards on the will

Ryan D. Tweney and Amy B. Wachholtz
Department of Psychology, Bowling Green State University, Bowling Green, OH 43403. tweney@bgnet.bgsu.edu amywach@bgnet.bgsu http://personal.bgsu.edu/~tweney

Abstract: Wegner's The Illusion of Conscious Will (2002) ignores an important aspect of the history of the concept: the determinism of William James and others. We argue that Edwards' formulation, and James' resolution of the resulting dilemma, are superior to Wegner's.

In 1754, Jonathan Edwards published his epochal Freedom of the Will. Edwards, a strict Calvinist predestination, sought to reconcile the omnipotence of a deity that directed and foresaw all future events, the "universal determining providence," with the need to regard people as moral agents who could be praised or blamed for their actions. Edwards rejected the idea of an autonomous causal chain (which he saw as inconsistent with divine omnipotence), and instead located human action within a deterministic network of unfolding events; a systemic model of the causation of behavior in which the force of the "apparent good" of an action was, like the force of gravity in planetary motion, the impetus for all behavior (Edwards 1754/1957). All ultimate causation was both divine and divinely foreordained, and the unfolding of human action was both part of a dynamic system and completely determined.

Edwards's book opened 150 years of debate in America about the nature and existence of "free will," an important debate in a liberalizing age that moved past the strict Calvinism of Edwards. Many of the opponents of determinism (e.g., Catherine Beecher and Rowland Hazard) appealed to the strong personal awareness of self-agency: Free will was self-evident, to be found in the consciousness of effort that accompanied all exertion of will. Such consciousness could not, so the argument went, be illusory, a claim that, as for Wegner, seemed wide open for psychological debunking.

The issue weighed especially heavily on William James, for whom all of science seemed to point to a strictly determined will. Yet this was an idea that led him into a stifling depression, one he mastered, as he tells us in an 1870 diary entry, by asserting his first great pragmatic formulation: "The first act of free will is to believe in free will" (H. James 1920, pp. 147). This characteristically Jamesian solution was a pragmatic stance that resolved his amotivational depression and eventuated in his later philosophical and psychological positions (James 1890). James's "solution" to Edward's dilemma effectively closed the debate begun in 1754, and since then free will has been only a minor topic in psychology.

Wegner tells us his readers nothing of this history; although it is directly relevant to many of the issues he raises. Thus, consider Edward's claim that "The will is as the greatest apparent good." (Edwards 1754/1957, p. 142). The phrasing was carefully explained: there is a seamless connection between the perception of an action as good at the moment and the carrying out of the action itself; each "is as" the other. Edwards used the example of a drunkard sitting before a drink; if the drunkard drinks it, it is because that action is the greatest good at the time. If the drunkard abstains, then abstention appears as the greatest good at the time. Either way, perception of the greatest good and the action are inseparable. Edwards rejected any notion of the will as an efficient cause, that is, as a mechnistic prior event, a motivational "cue ball" hitting an actionable "eight ball" (Tweney 1997). Human action—the perception of an action and its carrying out—were instead parts of a whole. The drunkard does not "freely choose" to drink. Instead, bad character and the proximal situation determine the drinking, just as were the character good, abstinence would prevail.

Consider now Wegner's definition of the experience of will: "Will is experienced as the result of self-perceived apparent mental causation" (pp. 65–66). This resembles Edwards's "The will is as the greatest apparent good," but it muddles the causal issues. Not that Wegner sets up his definition by referring first to a billiard ball notion of causality (see p. 64), as if that were the only kind of causation possible. This is not like Edwards's systemic determinism, in which the perception of the apparent good and the action are distinct but inseparable parts of a whole. The result is that Wegner fights, yet again, the old battle about the causal efficacy of the "feeling" of will. His book takes on the task of proving once and for all, that the feeling of will is illusory. This ignores the deep dilemma set by Edwards, as well as the elegance of the resolution profiored by James. Edwards needed a deterministic will because he sought a theocratic social order in which the "Elect"
Commentary/Wegner: *Précis of The illusion of conscious will*

Conscious Will (Wegner 2002) should not be lightly dismissed. It is, nevertheless, an affront to common sense. So it is equally important to outline the ways in which free will is not an illusion.

In what senses is conscious free will an illusion? First, it is an illusion in the sense that the causal role of any conscious experience in a "conscious mental process" can be said to be an illusion. In Velmans (1991a) I have suggested that a mental process might be conscious (a) in the sense that one is conscious of it, (b) in the sense that it results in a conscious experience, and (c) in the sense that conscious experience plays a causal role in that process. As Wegner shows, experienced will is a representation of what is going on in the mind/brain, making the mental processes represented by experienced will "conscious" in the sense that we are conscious of them (sense [a]). Preconscious decision making processes can also be said to become conscious once they result in a conscious free will experience (sense [b]). Wegner, however, gives many reasons to doubt that the experience of will actually governs the choices and decisions required for voluntary control (sense [c]), and I have given many additional reasons to doubt that conscious experiences govern the mental processes to which they most obviously relate in Velmans (1991a; 2000; 2002b; 2003b). In Velmans' experience of will case, conscious processes may represent them without governing them. We nevertheless feel that our conscious will determines our decisions and actions. That is the illusion.

Being representations of preconscious and unconscious mental processes, conscious experiences can also, occasionally, be misrepresentations, and Wegner provides various examples of misattributed volition (where people believe themselves to have willed an act that was determined by external forces or believe external forces to have determined acts that are actually carried out by themselves). That is a second sense in which experienced free will can be in illusion.

Such illusions of free will suggest that it may be causally epiphenomenal, which has threatening consequences for our moral and legal judgments, let alone our visions of our own agency. Consequently, many are concerned, as I am, to discern any overtones in which experienced will is not an illusion. According to him, "conscious will" is a feeling that informs us whether we, rather than an external agency, are the authors of acts, and helps us keep tally of what we are doing and what we have done (p. 328). This in turn helps establish a sense of who we are and gives us a sense of responsibility that leads to morality. I entirely agree, but only because this is a true story told from a first-person perspective, which doesn't, unfortunately, escape epiphenomenalism. Our conscious sense of "who we are," of "authorship," and of "responsibility" are as much experiences as are experiences of free will. And preconscious and unconscious processes determine our sense of self, authorship, and feeling of responsibility as much as they do our feeling of will. If from the perspective of brain science, experienced will is epiphenomenal, then from the perspective of brain science the same can be said of these other experiences. If one is to escape epiphenomenalism one has to do so another way.

As far as I can tell, a satisfactory account needs to make sense of how conscious experiences relate to their neural causes and correlates, and to the processes that they represent; it also needs to explain how first- and third-person accounts can be compatible, complementary, and mutually irreducible within a dual-aspect theory of mind (see Velmans 2000; 2002a; 2002b; 2003a; 2003b). Given BBS commentary space constraints, what follows is only a hint.

Note first that conscious experiences can be representations not just of our own minds, but also of our bodies and the surrounding physical world. In everyday life we behave as "naive realists." We habitually take the events that we experience to be the events that are actually taking place. Although sciences such as physics, biology, and psychology might represent the same events in very different ways, this approximation usually serves us well. When playing billiards, for example, it is safe to assume that the balls are smooth, spherical, coloured, and cause each other to move by me-

---

*Why conscious free will both is and isn't an illusion*

Max Velmans

Department of Psychology, Goldsmiths College, University of London,
London SE14 6NW, United Kingdom. m.velmans@gold.ac.uk

http://www.goldsmiths.ac.uk/departments/psychology/staff/velmans.html

Abstract: Wegner's analysis of the illusion of conscious will is close to my own account of how conscious experiences relate to brain processes. But our analyses differ somewhat on how conscious will is not an illusion. Wegner argues that once conscious will arises it enters causally into subsequent mental processing. I argue that while his causal story is accurate, it remains true that authenticates the action's owner as the self" (p.327). Free will is a positive illusion for Wegner — a signpost of reality, a mood enhancer, supplemented perhaps by a new kind of Zen-like resignation. This is a new predestination, perfectly appropriate to a psychologized society of isolated selves fantasizing control over their own fates; the illusion that because things seem to be so, then they are so. Do we have a "new Calvinism" here, one in which a psychological "Elect" can praise and blame those fooled by their own illusions of free will?

---

Dan Wegner has written a fine, insightful book that has genuinely gained some must prove for it. Thus, we could prove that there is free will by pointing to biofeedback studies in which people can learn (within limits) to control their own autonomic functions, such as heartbeat. In fact, there is no proof either way, since the construct of "free will" is not a psychological construct.

Finally, what of Wegner's moral claim? In the end, quoting James, Wegner brings in character as the warrant for praise and blame (p.324), dismissing free will as a potential psychological mechanism (but then why include all the experimental studies?), while reasserting the importance of the now-illusory emotion: "Conscious will is the somatic marker of personal authorship . . . that authenticates the action's owner as the self" (p.327). Free will is a positive illusion for Wegner — a signpost of reality, a mood enhancer, supplemented perhaps by a new kind of Zen-like resignation. This is a new predestination, perfectly appropriate to a psychologized society of isolated selves fantasizing control over their own fates; the illusion that because things seem to be so, then they are so. Do we have a "new Calvinism" here, one in which a psychological "Elect" can praise and blame those fooled by their own illusions of free will?
Commentary/Wegner: Précis of The illusion of conscious will

The short- and long-term consequences of believing an illusion

Michael E. Young

Department of Psychology, Southern Illinois University at Carbondale, Carbondale, IL 62901-6502; meyoung@siu.edu

http://www.siu.edu/~psycho/bcs/young.html

Abstract: The experience of free will has causal consequences, albeit not immediate ones. Although Wegner recognizes this, his model failed to incorporate this causal link. Is this experience central to “what makes us human”? A broad acceptance of Wegner’s claim that free will is illusory has significant societal and religious consequences, therefore the threshold of evidence needs to be correspondingly high.

Wegner (2002) has produced a thought-provoking idea — conscious will is the by-product of an error-prone causal learning system — and an entertaining recitation of a number of phenomena supporting his thesis. Some of the empirical data lack the experiential rigor of today, but there is indeed ample support for the existence of illusions of conscious will. Although there is much to like about Wegner’s book, there are two issues on which I will focus my attentions — a missing link in his model, and the societal and religious implications of a broad acceptance of his thesis.

In his consideration of the experience of conscious will, Wegner proposes a model (summarized in Fig. 3.1 of his book) in which the (conscious) intention of doing an action consistently precedes the action only because they both arise from the same source (an unspecified unconscious thought) and the intention occurs earlier than the action. Thus, conscious will is an illusion because the intention has no direct causal link to the action that it purportedly causes. The intention appears to have caused the action because it satisfies at least three of Hume’s cues-to-causality: temporal priority, temporal contiguity and covariation, and, perhaps, spatial contiguity (if one can locate thought in space). Causal learning and inference is now known to be quite complex, involving interactions of various types and an interaction between time and contingency (e.g., Cheng & Novick 1991; Shanks et al. 1998; Young et al. 2000a; 2000b). Be that as it may, any induction of causation is subject to error and thus may be illusory.

Throughout my reading of his book, I was persistently bothered by the notion that this illusion might be unnecessary — why would our species need to feel like we intended our actions when the mere consequence of our actions could serve the purpose of selecting the appropriate responses for various situations in the future (à la Skinner)? Does the illusion have any benefit to us as a species? Apparently, Wegner was bothered by the same question as revealed in his final chapter, “The Mind’s Compass.” He likens the will to an authorship emotion, a feeling of responsibility for our actions, and proposes that this emotion affects our future behavior. Thus, experiencing the illusion of free will does have consequences. This supposition suggests that there is a missing link in Figure 3.1 from the experience of conscious wills backward to the production of future unconscious causes of thoughts and actions. The experience of conscious will may indeed have an effect on our species if it serves to alter our future mental states, thus raising will from an epiphenomenon to a true cause of behavior. But, its causal effects are distal — the intention did not cause the action that immediately followed it (see Fig. 3.1), but the experience of the efficacy of the intention would affect future actions by altering the likelihood of future unconscious thoughts.

What then are the consequences of not experiencing free will? Wegner suggests that we would be amoral, psychologically unhealthy individuals, prone to depression, anxiety, and a general feeling of helplessness. Perhaps, he fails to consider the behavior of other species — do other species experience free will and if not, are they amoral and psychologically unhealthy? Perhaps. Without the ability to disable this experience in ourselves or others, answering such questions belongs in the realm of speculation. Although the speculations raised in the final chapter of the book are thought provoking, they are not definitive because of the difficulty (impossibility?) of measuring mental experiences (Uttal 1998). Does a criminal truly feel no responsibility for his actions or is he simply better at suppressing the normal visible responses associated with feelings of guilt?

The potential costs of an acceptance of Wegner’s thesis are disturbing as he readily acknowledges in his final chapter. The results could be abdication of personal responsibility and an undermining of traditional religious morality. If my actions are merely the byproduct of my environment, then the environment is to blame when I fail, commit a crime, or sin. These attitudes seem increasingly prevalent in today’s society, especially with an appeal to biological (genetic) determinism. A corollary not as readily accepted, however, is that the environment should also receive credit when I succeed, do a good deed, or follow God’s laws. This asymmetry of attitudes about responsibility is so prevalent that it has its own name — the fundamental attribution error.

The core tenets of most religious schools of thought actively dismiss a blaming or crediting of the environment; rather, they advocate for personal responsibility for the consequences of one’s actions. Although faiths differ in many ways, God is typically viewed as one who metes out justice for one’s sins and rewards for one’s good deeds, and not as an arbitrary force of nature. Thus, any school of thought that advocates for an absence of free will (e.g., Skinner’s behaviorism or Wegner’s illusory free will) works in opposition to most contemporary religious thought. Is this a battle that Wegner wishes to fight?

Wegner never effectively addresses the consequences of a widespread acceptance of his thesis. He seems to say that although free will is an illusion, it is one that we should all maintain because “the illusion makes us human” (p. 342). Fortunately, the data that he cites, although representing a wide array of situations under which we do experience illusory causation, can only prompt an inquisition, not a deduction, of his thesis. It is still possible that many of my experiences of conscious will are not illusory; and that is an “illusion” that I will continue to comfortably harbor.
Conscious will and agent causation

G. E. Zuriff
Department of Psychology, Wheaton College, Norton, MA 02766.
Gerald.Zuriff@verizon.net

Abstract: Wegner (2002) fails to (1) distinguish conscious will and voluntariness; (2) account for everyday willed acts; and (3) individuate thoughts and acts. Wegner incorrectly implies that (4) we experience acts as willed only when they are caused by unwillful thoughts; (5) thoughts are never true causes of actions; and (6) we experience ourselves as first performing mental acts which then cause our intentional actions.

1. Although Wegner (2002) attempts to establish psychology on a scientific footing, his conceptualization of the causation of behavior is based on a non-operationalized and subjective distinction between voluntary behavior and behavior experienced as brought about by conscious will. In the experiments Wegner reviews to support his thesis, it is never clear how subjects interpret the instructions to rate how much they intended to perform a particular action. Are they measuring the degree of conscious will experienced? Or reporting the extent they experienced the action as theirs? Or how much they intended the action? What exactly are they reporting?

2. Wegner’s distinction between voluntary and conscious will is also blurred in everyday action. When I walk to work every morning after breakfast, I may have no prior thoughts or plans about my action. Yet it is clear to me that my walking is voluntary, intentional, and my action. If on a particular occasion my act was rude or criminal, I am clearly morally and legally responsible for it. To be sure, there may be times when it is difficult for me to go off to work and I might say it took an “act of conscious will,” but I do not experience myself on the other more routine occasions as any less an agent. I conclude that, contrary to Wegner, I experience my walking as an intentional willed action even though I did not observe a regular correlation between prior thoughts and subsequent action.

Moreover, not only is my walking to work experienced as intentional and willed, but so is every step I take, as well. Does Wegner require that I observe a correlated thought prior to each step in order for me to have this experience? Of course, each step is itself segmented into smaller intentional and willed units (e.g., lifting my leg, bringing it forward). Does each segment require correlations with prior thoughts for me to infer that I have caused it?

4. Not only are my overt actions, like walking, experienced as willed and intentional, but so are my mental activities, such as trying to recall a name, or planning my day. However, in Wegner’s account, in order for me to have this experience of consciously willing a mental action, M, I must have inferred its causation by experiencing a correlation between it and some prior mental event, P. However, P is experienced as consciously willed, it too requires a prior correlated mental event, Q, and we are trapped in an infinite regress, with each consciously willed mental event preceded by yet another consciously willed mental event. To escape, we must allow that one mental event in the illusory causal sequence is not experienced as an intentional willed act. It follows from Wegner’s thesis, that for me to experience a mental action as intentional, I must experience it as ultimately caused by an unintended unwillful thought. This is both counterintuitive and contrary to our experience of ourselves as agents.

5. Wegner argues that the experience of thoughts and plans causing actions is an illusion. However, there are many instances when our mental activities in fact cause our actions. Consider what happens when I consult my shopping list before buying a cereal. Unless Wegner is wedded to an outmoded notion of causation requiring temporal and spatial contiguity (i.e., the billiard ball model he is fond of), undoubtedly, the list, functioning as a stimulus, is a cause of my response. This type of causation is not different from saying that a reinforcement that occurred several days ago may be a cause for my response emitted today even though spatial and temporal contiguity is lacking. (This type of causation does not of course, preclude the existence of an underlying causal chain of neuro-physiological temporally and spatially contiguous events.) Suppose, now, I have memorized the list and it appears not on a piece of paper, but rather in my memory. When I recall the list and make my cereal selection, the mental list, serving as an internal stimulus, is a cause of my response. Thus, our actions are often truly caused by internal mental events like planning, recalling, calculating, and reasoning, with no illusions involved.

6. Ultimately Wegner’s concept of mind is based on a simple but flawed model. He assumes a self, independent of thoughts and actions, which experiences thoughts and actions distinct from itself. When the self notices a correlation between thoughts and actions, it infers that it intentionally caused the actions. This model in which the self, thoughts, and actions are logically separate, was analyzed and, I thought, demolished a long time ago by Gilbert Ryle (1949). In everyday life we know what it is like to do A by first doing B. For example, we turn on the TV by first pressing a button, and we lift the box by first pushing down on the lever. But we do not ordinarily experience ourselves as moving our arm by first doing an act of conscious willing, or anything else for that matter. In most cases we experience moving our arm as what Danz’s (1963) called a “basic action,” that is, we simply move it as the act of an agent. When we perform basic actions, we experience them as willed, intentional, and ours. Thus, in contrast to Wegner’s model, this alternative model suggests that the inter-dependent cluster of agency experiences – will, intention, self, basic acts – arise simultaneously, both conceptually and developmentally (Zuriff 1975; 1985, Ch. 9). There is no independent self that makes inferences and experiences itself as causing movements by first having thoughts.

ACKNOWLEDGMENT
The help of Hakadosh Barachu is gratefully acknowledged.

Author’s Response

Frequently asked questions about conscious will

Daniel M. Wegner
Department of Psychology, Harvard University, Cambridge, MA 02138.
wegner@wjh.harvard.edu http://www.wjh.harvard.edu/~wegner/

Abstract: The commentators’ responses to The Illusion of Conscious Will reveal a healthy range of opinions – pro, con, and occasionally stray. Common concerns and issues are summarized here in terms of 11 “frequently asked questions,” which often center on the theme of how the experience of conscious will supports the creation of the self as author of action.

In the course of giving talks on conscious will, I have found that the question period after the talk yields spirited interchanges, not to mention the occasional rubnurn. Like the commentators on The Illusion of Conscious Will (ICW) (Wegner 2002), audiences can be polarized. One tactic that I have found useful on these occasions is to summarize my talk as follows:

What you may have heard me say:
• Cognition does not cause action.
• Planning does not influence action.
• There is no intention and no responsibility.
• Life as we know it on earth is now over.
Response/Wegner: Précis of The illusion of conscious will

What I was hoping to say:

• Conscious will is based on interpreting one’s thought as causing one’s action.
• The experience of will comes and goes in accord with principles governing that interpretive mechanism and not in accord with a causal link between thought and action.
• The experience of conscious will thus is not direct evidence of a causal relation between thought and action.

This disclaimer suggests that people often read much more into ICW than is there. On studying the comments of this excellent field of commentators, I find that something like this has happened again. A variety of intriguing issues have been raised, several of which suggest important amendments to ICW, but some of the commentaries involve more being attributed to ICW than its pages actually held. Rather than trying to give the talk all over again, I propose to guide us through the question period in a way that will allow everyone out of the auditorium in time for the wine and cheese. To do that, I have organized what I hear people saying into a FAQ about Conscious Will.

As a backdrop for these questions, here is a Table with a very rough summary of what I think the commentators were saying. Their remarks can be sorted along two axes: whether they appear to agree or disagree with the main thesis of the book, and how they propose to contribute to the discussion.

This straw poll shows a field of commentators quite evenly divided — indicating that the topic is alive and worthy of continued consideration. ICW did not settle the controversy (alas), but it also did not merely pave over the sarcophagus of a question long decided. The ranks of commentators on each side are not too surprising. It is easy to see how Dennett would agree with ICW — despite the title of his latest book, Freedom Evolves (Dennett 2003a) — for example, and the positive votes of Glymour, Itô, Kirsch & Lynn, Mandler, Metzinger, Pylyshyn, and Velman make sense as well in light of their prior work. In various ways, these commentators have been pressing themes like those rendered in ICW for some time.

On the nay-saying side, Kihlstrom is especially fervent, representing the role of conscious will that appeared in his book as the end of life on earth. But there it is in black and white. Such shrill invective must have a motor, an emotional energy and conviction behind their responses is reminiscent of the strong words voiced by Nahinias (2002) in an earlier review of ICW:

Quoting: “If it isn’t literally true that my wanting is causally responsible for my reaching... and my believing is causally responsible for my saying... then practically everything I believe about anything is false and it’s the end of the world” ([Fodor] 1990, p. 196). We philosophers should keep our guards up against any blow that would be the end of the world.

Whew. You thought I was kidding that someone might see the book as the end of life on earth. But there it is in black and white. Such shrill invective must have a motor, an emo-

Table 1. Themes of the Commentaries

<table>
<thead>
<tr>
<th>Conscious will is an illusion, but...</th>
<th>Dennett, Glymour, Itô, Morton, Raz &amp; Norman, Velman, Young</th>
<th>Jack &amp; Robbins, Metzinger, Pylyshyn</th>
<th>Kirsch &amp; Lynn, Mandler</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conscious will is not an illusion, and...</td>
<td>Sterneberg</td>
<td>Ainslie, Handcastle, Heyman, Krueger, Schultz, Sebanz &amp; Frith</td>
<td>Tveney &amp; Wachholtz</td>
</tr>
<tr>
<td>Can't decide, but...</td>
<td>Panksepp</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

... someone else has already thought of this. 

... here's a good scolding.
tional basis that drives the rhetoric and motivates some commentators to find not one error in the book, or a few, or even dozens, but instead to pronounce the entire work "an unwarrantable impertinence" (Kihlstrom). Several of the commentators express similar views (Hardcastle, Bogen, Sternberg, Zuriff), albeit somewhat less breathlessly. Where is this passion coming from?

I think it is the personal experience of conscious will that can make ICW hard to appreciate. Every day at innumerable junctures, we think of doing a thing and then do it. We think of getting a cup of coffee and do it; we think of checking our e-mail and do it; we think of taking a book back to the library and do it. These special instances of conscious will stand out, moments radiant in memory whenever we reflect on how it is that our actions occur. These experiences illuminate the guiding intuition that overwhelms our judgment. A fitting remark was made by Anais Nin: We don't see things as they are; we see them as we are.

It is difficult to savor this intuition of will and at the same time appreciate the unconscious mental and brain processes that create the intuition. The sense of conscious will seems to clash with its own causal underpinnings, contradicted as it were by the very events and processes that bring it into being. Curiously, the conflict between feelings and explanations of those feelings is not as strong in other niches of the mind. Not too many people would complain, for example, that they could no longer feel joy, or perhaps trust their feeling of joy, if science could deliver to them a good explanation of the unconscious circuits and circumstances that gave them the feeling of joy. Joy is not ruined by its explanation. Yet trying to discern the causes of conscious will seems to tamper with the meaning of the experience.

This is true because the experience of conscious will is involved in the creation of the self. The feeling of doing establishes a "doer," not only authenticating the self but constructing the self from what was previously thin air (Wegner, in press). Those actions for which we feel no conscious will—such as the absent-minded noshing we do when a bowl of snacks is at hand—are actions that need no author. Such invisible actions arising without consciousness do not require us to be the ones who did them. They just seem to happen, and oh, by the way, they happened to us. Consciously willed actions, on the other hand, give the mind the opportunity to identify itself as author. Something was done because it felt like it was done, so there must have been an "I" who did it.

Like the commentators who defend the experience of conscious will, most of us see the experience of conscious will as necessary for our personal survival. This intuition is built into the human mechanism, preordaining a discomfort that ranges from uneasiness to pure existential dread whenever we ponder the possibility that conscious will is an illusion. Adopting the premise that conscious will is an illusion, then, is extraordinarily difficult, a move that seems to deny the ever-present reality of our selves. As Jack & Robbins note, "a genuine illusion remains compelling even when the subject knows their experience is misleading." One cannot really just "stop believing" that one has a self, no matter how useful this would be to one's theorizing, so the intuition constantly clashes with the kinds of things one must think in order to understand how the self is created.

But of course, each self must have been created sometime, somewhere, somehow. Short of imagining eternal souls, most of us recognize that there is a point in each hu-

Response/Wegner. Précis of The illusion of conscious will

One of the basic assumptions of ICW is that the primary manifestation of conscious will is the person's reportable experience of consciously willing specific actions. This is the center of the intuitive world of will, but a number of commentators suggest that this emphasis is in error. Ainslie describes the experience of will as only a facet of the phenomenon, noting that there are other important ways of understanding motivated and self-controlled behavior. Hardcastle views the sensation of will as entirely irrelevant to the question of whether conscious will is an illusion ("the sensation of will isn't the will itself"). Jack & Robbins suggest that limiting the idea of will to the experience of will prejudices the issues, and maintain without explanation that "it is perfectly coherent to claim that you have consciously willed something without having an experience of doing so."

The experience of conscious will is, of course, the basis of the intuition we all have that we cause our actions. Without this intuition, we might find it relatively painless to have free will extracted from our conceptions of behavior causation. In everyday discourse about action, though, the experience is essential: If you say you consciously willed leaving the ice cream out of the freezer to melt on the kitchen counter, you will get far more flack from any ice cream lovers who live with you than if you say you did not consciously will it. If you do not admit consciously willing leaving it out, it will not even matter if you say that you thought of leaving it out, or if you admit that it was indeed left out. You don't think you did it.

Legal decisions similarly rest on reports of the feeling. If a shooter claims not consciously willing the shot that killed a person, a jury—a least right now, in America—will be at pains to convict the shooter of murder. Perhaps it was manslaughter (Demno 2002). The fact that animals (and plants and computers and babies and lots of other agents) simply cannot report consciously willing their actions, in turn, imposes a massive obstruction on anyone wishing to suppose that one of these agents indeed consciously willed
Response/Wegner: Précis of The illusion of conscious will

any act (cf. Macphail 1998). If you cannot self-report an experience of conscious will, the conscious willing of the action may not have happened. The emphasis of ICW on the experience of conscious will is the same emphasis we all place on reports of the experience in every phase of life. Feeling is the essential ingredient of conscious will, not an add-on.

Certainly, there are many ways of defining the will that leave out the experience. At its most general, willful behavior can be said to occur whenever there is evidence that information input to a system caused a change in the subsequent behavior of the system (Garver & Scheier 1998; Kennedy 1992; Miller et al. 1960; Powers 1990; Wiener 1948). This fundamental form of will has never been at issue in ICW — indeed, I am convinced that control systems are good models of the architecture of human behavior production (Vallacher & Wegner 1985; Wegner & Bargh 1998). "Thoughts must cause actions in this sense. This is the empirical will as defined in ICW. It is only when we add the experience of conscious will to the system that everything becomes murky. Heyman reports that ICW overlooks "the objective basis for the sensation." It does so because the book simply assumes intelligent goal-seeking behavior on the part of humans. The experience of such behavior is the issue.

The experience of conscious will often results in self-reports. The person volunteers something to the effect of “I thought of getting up for a soda and then I did.” Asked whether he or she consciously willed getting up for the soda, the person would likely say yes. A basic point of ICW is that the person’s self-report of this experience is not a di- somatic report. The feeling is the key to conscious will, not the validity of conscious will in behavior explanation, it seems to slip into the ruts. This seems to have happened to several commentators in the interpretation of the illusion. Some commentators deftly replaced my notion of the experience of conscious will the idea of an “act of will” — a temporally distinct mental event whose place in the causation of action they saw at issue. Unlike ICW’s notions of empirical will and phenomenal will, this notion captures something more like a spark of causation that sets off action. This interpretation was pursued by Krueger in a detailed causal analysis, and was similarly developed by Sternberg with the suggestion that human behavior “is part of a complex chain of events in which conscious will does not come at the beginning of the chain.” The ICW analysis contains nothing that corresponds specifically to the causal spark envisioned by these commentators. Calling conscious will an illusion sounds like a call to snuff out that spark — to douse the flame of our selves.

What then is the illusion? The illusion I hoped to explore in ICW was the illusion of a self-knowing causal mechanism. Let me explain. Most of us would regard as absurd the idea that causal events should know that they have occurred. When a tree topples over in a forest and falls into a pond, for example, we are fully content to assume that no one knows. The tree doesn’t know, the pond doesn’t know, and the tree-falling-into-the-pond doesn’t know. I bring up this entirely bizarre way of speaking about physical causation to contrast this case with the puzzling events of mental causation. When a person thinks of diving into a pond and then does so (and feels this was consciously willed), we normally accept that the causal mechanism underlying this event knows itself! The person’s report on how it happened (“I consciously willed diving into the pond”) is taken as a privileged communication from somewhere way down in the causation factory indicating exactly what happened. This simply cannot be the case. If it does not make sense for physical causation, it also should not make sense for mental causation.

We only understand mental causation in ourselves by virtue of an authorship processing system that examines a variety of indicators to determine whether this particular action was one we consciously willed (Wegner & Sparrow, in press; Wegner et al. in press). The authorship processing system stands outside the processes that cause the action itself, laboring in parallel with it to generate feelings of doing that inform us of an estimate, based on available information, of who did the action. These estimates are accompanied by associated experiences of conscious will. The authorship processing mechanism gives rise to experiences of conscious will that compel us to believe that we cause our actions.

Ultimately, the illusion of conscious will comes down to an issue of the validity of self-reports of mental processes. Nisbett and Wilson (1977) pointed out that mental processes do not explain themselves — for example, that the process of choice does not necessarily “know” what produced a particular choice, or that the process of addition does not “know” how the computation was carried out. In this view, the products of mental processes may be knowable, but the mental processes themselves are not “self-luminous,” the wonderful term for our mind’s flattering self-portrait coined by Ryle (1949). It is a mistake to think that

R3. What does it mean to call conscious will an illusion?

Before ICW was written, I had alternate titles for it: The Construction of Conscious Will, The Experience of Conscious Will, The Fabrication of Conscious Agency, and so on. Not far from these is Heyman’s suggestion: The Sense of Conscious Will. All of these titles lead to much the same point as ICW, but without the word illusion. I was even admonished by Dan Dennett prior to publication that illusion was a fighting word. Several commentators agree that illusion is loaded, and a number disagree on just what it is that ICW describes as illusory.

One common response to the word illusion is to see it as a challenge to the causal properties of mental events — the idea that thought causes action, or that plans influence action. There are some very deep grooves in the road running from determinism to free will, ways of thinking about things that have been so well trodden that it is easy for any vehicle to slip into the ruts. This seems to have happened to several commentators in the interpretation of the illusion. Some commentators deftly replaced my notion of the experience of conscious will the idea of an “act of will” — a temporally distinct mental event whose place in the causation of action they saw at issue. Unlike ICW’s notions of empirical will and phenomenal will, this notion captures something more like a spark of causation that sets off action. This interpretation was pursued by Krueger in a detailed causal analysis, and was similarly developed by Sternberg with the suggestion that human behavior “is part of a complex chain of events in which conscious will does not come at the beginning of the chain.” The ICW analysis contains nothing that corresponds specifically to the causal spark envisioned by these commentators. Calling conscious will an illusion sounds like a call to snuff out that spark — to douse the flame of our selves.

What then is the illusion? The illusion I hoped to explore in ICW was the illusion of a self-knowing causal mechanism. Let me explain. Most of us would regard as absurd the idea that causal events should know that they have occurred. When a tree topples over in a forest and falls into a pond, for example, we are fully content to assume that no one knows. The tree doesn’t know, the pond doesn’t know, and the tree-falling-into-the-pond doesn’t know. I bring up this entirely bizarre way of speaking about physical causation to contrast this case with the puzzling events of mental causation. When a person thinks of diving into a pond and then does so (and feels this was consciously willed), we normally accept that the causal mechanism underlying this event knows itself! The person’s report on how it happened (“I consciously willed diving into the pond”) is taken as a privileged communication from somewhere way down in the causation factory indicating exactly what happened. This simply cannot be the case. If it does not make sense for physical causation, it also should not make sense for mental causation.

We only understand mental causation in ourselves by virtue of an authorship processing system that examines a variety of indicators to determine whether this particular action was one we consciously willed (Wegner & Sparrow, in press; Wegner et al. in press). The authorship processing system stands outside the processes that cause the action itself, laboring in parallel with it to generate feelings of doing that inform us of an estimate, based on available information, of who did the action. These estimates are accompanied by associated experiences of conscious will. The authorship processing mechanism gives rise to experiences of conscious will that compel us to believe that we cause our actions.

Ultimately, the illusion of conscious will comes down to an issue of the validity of self-reports of mental processes. Nisbett and Wilson (1977) pointed out that mental processes do not explain themselves — for example, that the process of choice does not necessarily “know” what produced a particular choice, or that the process of addition does not “know” how the computation was carried out. In this view, the products of mental processes may be knowable, but the mental processes themselves are not “self-luminous,” the wonderful term for our mind’s flattering self-portrait coined by Ryle (1949). It is a mistake to think that
reported experiences of mental processes are valid indicators of the causal sequences underlying those processes. As noted by Pylyshyn: “Even though our causal mental process may go through a sequence that corresponds to the sequence that we experience, it does not in general proceed that way for the reasons suggested by the conscious experience” (his emphasis). The illusion of conscious will is the belief that we are intrinsically informed of how our minds cause our actions by the fact that we have an experience of the causation that occurs in our minds.

R4. What conclusion should be drawn when the feeling of conscious will is mistaken?

One of the main projects of ICW is to catalog cases in which the experience of consciously willing an action is at variance from the observed causal sequence whereby the action occurs. A person feels that she has consciously intended to point at a duck, for example, but it was arranged in advance that someone else holding her hand actually would move it to point to the duck (e.g., Wegner & Wheatley 1999). There are many cases like this one in which conscious will is present but verifiably voluntary action is not present, and there are also cases in which verifiably voluntary action occurs without benefit of the feeling of conscious will. Several of the commentators (Hardcastle, Kihlstrom, Metzinger, Schultz et al., Tweney & Wachholtz) argue that the listing of exceptions to the efficacy of conscious will does not invalidate conscious will overall, and that the book’s project is therefore in error.

This doesn’t make sense to me. If someone has a theory that the earth orbits around the sun, for example, and it turns out that on Wednesday, February 4, 2004, it briefly orbited around a Wal-Mart store in Duluth, I’d say that pretty much shoots the sun theory. Exceptions do not prove rules, we all know – they invalidate them. If the experience of conscious will is indeed connected in any but the most capricious way to the causal sequence whereby actions occur, it should not be mostly right but sometimes wrong. It should be perfect. If the feeling of conscious will is intrinsically right, informed somehow by the fact that it is the cause of an action, it should lock on to that causal relationship and always reflect it correctly. The minor exceptions these commentators would allow in the accuracy of reports of conscious will – such as hypnosis or spirit possession or the automatisms – are not minor at all. They rend the fabric of conscious will from one end to the other.

This logic requires that we draw an important inference. From “The feeling of conscious will can be mistaken,” we must conclude that “the feeling of conscious will is never correct.” I agree that on its face, this seems extreme – Metzinger calls it a “non sequitur.” But there is a deeply important reason for making this strong inference: Imagine the horrible kludge we would have to create to accommodate a partly valid experience of conscious will in any mental system we might envision underlying human behavior. In essence, we would need to create a double system – one to produce willed acts in which the feeling was authentic, and another to produce acts in which the feeling is disconnected from the action and did not reflect how the action has occurred. And what is more, this engineering nightmare would also require the installation of a higher-level super system that would need to determine when each of the sys-
and action is not the same as saying that mind does not have a causal relationship to action. It could, and in fact we all should be fairly certain that it does. I hoped that in defining the empirical will as this actual causal relationship between mind and action, and the phenomenal will as the experience of this relationship, I had made things sufficiently clear that no one would fall into the trap of thinking that these were the same thing. Yet here we have a number of commentators who have made exactly this mistake.

Mandler reports that “Wegner threatens to throw the baby out with the bathwater when he implies that mental events can never be causal agents for thought and action.” I never said never. Jack & Robbins pounce gleefully on what they interpret as my dramatic reversal: “Wegner appears to have realized that he can’t defend this thesis, conceding in a later publication that we can construct a scientific account of consciously willed action.” In a related vein, Tweney & Wachholtz note “Thus, we could prove that there is free will by pointing to biofeedback studies in which people can learn (within limits) to control their own autonomic functions, such as heartbeat.” In saying this, they reveal their assumption that the actual relationship between thought and action is at issue here. It is not. Sometimes I wonder what book these folks were reading.

ICW is a theory of the experience of will, not a theory of the relationship between thought and action. It is entirely possible that experimental evidence can be assembled to establish to my and everyone else’s satisfaction that a conscious (reportable) thought is a cause of an action in a given setting. I have found it useful to assume that thought does cause action in many, many ways (Vallacher & Wegner 1985; Wegner 1989; Wegner & Pennebaker 1993; Wegner & Vallacher 1977) and am a bit surprised that even this minority of the commentators would mistake my meaning. The fact is, though, that no one has fully demonstrated the causal necessity of conscious thought in a scientific way with anything even remotely near the impact of the demonstrations we give ourselves all day when we think of things and do them. As long as we as humans mistake the feeling for the causing, we are scientists won’t gain good evidence for the causal.

Several investigators have been trying to examine this issue scientifically. Velmans (1991a) approached the problem by a process of elimination, examining whether consciousness is an essential ingredient of cognitive processes or behavior-production systems. His early conclusion was that it may not be required, and in his current commentary he develops that idea a bit more. Bargh (e.g., Bargh & Ferguson 2000; Bargh et al. 2001) has pursued a similar agenda by examining whether complex voluntary actions might be produced without benefit of consciousness, also suggesting the conclusion that consciousness might not be necessary. In contrast, Baars (e.g., 1988; 2002) has focused on the cases when consciousness seems to be essential, assembling examples when consciousness may contribute uniquely to the course of thought and action. There is some promise in studying whether conscious processes activate brain mechanisms that are not activated without consciousness – mechanisms that then show a causal influence on behavior – and this may be another helpful way of furthering this line of inquiry (e.g., Jack & Shallice 2001). In sum, the question of whether there is an empirical will, as noted in ICW, is an empirical question – not one to be settled by my presumption that there is no such thing, or by the presumption of others that I have been presuming too much.

R6. Is conscious will that occurs before action more causal than conscious will that occurs during action?

A number of commentators (Heyman, Krueger, Schultz et al.) brought up the distinction between prior intentions and intention in action (Bratman 1984; Searle 1983), hoping that this might solve the problem at hand. They noted that people think about doing things sometimes well in advance of acting, and that these conscious willings might often be causal even when the feeling of will occurring just as the action is running off might not be there in time to be causal. This comment suggests that these commentators saw the point of ICW as being the same as the one raised by Libet’s classic studies (Libet 1985; Libet et al. 1982) on conscious will in finger movements. His research suggests that the experience of conscious will at the moment of deciding to move may precede the action, but that the experience of willing nonetheless follows brain events associated with the choice by several hundred milliseconds.

The timing of the experience of will with regard to action brings up some useful observations, but it is not particularly telling regarding the main thesis of ICW. The idea of ICW is not dependent on Libet’s findings – although it isn’t damaged by them either. This is because ICW (again, for those who’ve not been listening) is not about whether thought causes action. It is about whether the experience of conscious will reflects such causation. So, just because an experience of conscious will happens well in advance of brain events leading to an action (rather than some milliseconds afterward) does not mean that the experience is any more direct an indication of the causal process whereby that action was produced (than one occurring later on). The experience of conscious will is a mental reconstruction of the causal sequence whether it happens well before the action, well afterwards, or right on time.

Experiences of conscious will occurring for an action right now might, however, have causal implications for subsequent actions or events. This is the point elaborated nicely by Young. Indeed, this downstream causality is what keeps the experience of conscious will from being the sort of useless, epiphenomenal appendage that some envision when they think of determinism. The experience of conscious will for an action at one time is an event with repercussions later on: the person can report it, can remember it, can have subsequent thoughts based on it, and so on. This means that such an experience can have influences galore for events that follow the experience. Feeling that one has consciously willed an action that hurt a friend, for example, would seem to create conditions that could lead to later behavioral changes: apologies to the friend, adjustments of the behavior, and so on. Memory for what one seems to have consciously willed in the past can influence the direction of subsequent behavior in the future. Experiences of conscious will are not erased forever from candidacy for behavior causation by their lack of validity as explanations of the actions to which they refer.

R7. Where does the theory lead empirically?

If you have been paying close attention, you will know that nobody actually asked this question. In a curious way, this set of commentaries, as well as my response, involves a
There are other research paths that ICW suggests, new ways of understanding how people keep track of agency, develop selves, and perceive their own actions against the backdrop of the actions of others and events in the world. People rapidly process information indicating their degree of authorship in the production of actions, thoughts, and experiences of responsibility and personal authorship, says Preston & Wegner (2003). Although it seems that ICW has turned out to be of some philosophical interest, it was written with science in mind.

The development of new research directions might help to assuage the concerns of some commentators. More scientists than philosophers, they anguish over the big muddy issues of the book has stirred up and some hope for better days. Mandler says “Maybe it would be best to forget about the problem of will altogether,” and Panksepp offers the hopeless hope that “future generations also need to endlessly debate these scientifically unfathomed and perhaps unfathomable neuropsychological issues.” Maybe the better proposal for all of us is “back to the lab!”

R8. Who came up with this idea anyway?

Whom should we credit the authorship of this theory of authorship? Since I’m defending the damn thing at some length, I’m feeling kind of authorly right now – but I am all too aware of the many contributing strains of thought to claim that the apparent mental causation theory is new under the sun. Several commentators helped to remind me of my secondary role by serving up explicit reminders of prior theories. Kirsch & Lynn seem to suggest that I lifted the theory from their unpublished submission to a journal I was editing. This view could only follow from an indulgently sympathetic reading of their work (e.g., Kirsch & Lynn 1999b), along with a studied ignorance of the rest of psychology. Indeed, Mandler reports that ICW “rehearses an argument that has dominated scientific psychology for about a century,” and goes on to suggest that Westcott (1977) also presaged the theory. Tweney & Wachholtz reach back yet earlier to attribute the idea to Jonathan Edwards (1754/1957) (although I fear on reviewing their analysis that I would never have found it there on my own reading). Velmans comments in passing on the “convergence between [Wegner’s] views . . . and my own conclusions.” Given this full range of attributions, I turned to reflect on my own theory of the origin of the idea.

I have been keeping my eye on a far different set of cothinkers. Most centrally, the early clear harbinger of this theory is David Hume (1739/1888), although I had not studied Hume’s thinking on this until I was in the middle of the project. His view that experiencing will involves perceiving causation within one’s own mind is exactly what I am talking about. Another remarkably prescient spokesperson for the idea that will is a causal inference was Michotte (1963). After these theorists, the lineage becomes more crowded. Although I did not realize it until after ICW was complete, Nisbett and Wilson (1977) can be read to say several of the essential ideas of apparent mental causation.

In 1996, as I began work on ICW, I encountered the writ-
ings of Brown (1989), Hoffman (1986), and Spence (1996), each of whom nailed some major part of the idea and made me feel that the theory had been scooped. In other ways, works by Bem (1972), Dennett (1992), Gazzaniga (1983), Harnad (1982), Langer (1975), Frith and Done (1989), Gopnik (1993), Spans (1986b), and yes, Kirsch and Lynn (1999b) also captured facets of the idea and stayed in my mind as important influences. The work of Bargh on automaticity (Bargh 1997; Bargh & Chartrand 1999) served, too, as a major reminder that some of the richest and most complicated human behaviors are caused by mechanisms that do not require conscious will.

When Thalia Wheatley and I published the first paper on the theory of apparent mental causation (Wegner & Wheatley 1999), many of these key ideas had already come together to influence us. We added to the mix our own naming of the variables governing inferences of causation from thought to action (consistency, priority, and exclusivity), although these had been recognized in various ways in the causal attribution literature (Gilbert 1995; Heider 1958; Jones et al. 1972; Kelley 1972; Michotte 1963; Nisbett & Ross 1980; Young 1995). My own best guess about the original contribution of ICW is in this area – the proposition that an internal causal analysis leads to the feeling of will. This paper then formed a chapter of ICW, and the book was designed to illustrate how this idea might have empirical consequences. Just as ICW was being published, I was surprised to encounter independent theoretical statements similar to apparent mental causation by Claxton (1999) and by Thompson et al. (1998). Notwithstanding Mandler’s judgment that ICW echoes the dominant view of the past century of scientific psychology, it also is an idea whose time is now.

It seems clear that determinism has been with us for a long time, as has the feeling of conscious will. This particular way of trying to put them together has dawned on a lot of people. Now, as it happens, thinking of things and keeping track of the fact that you thought of them are two quite different mental processes – the first a matter of mental causation, the second a matter of apparent mental causation and authorship processing. If I promised you that I knew I was the author of every idea I wrote, I would be claiming that my mental processes somehow knew their own authorship. I would be claiming, in short, that my conscious will created these ideas. With Ryle (1949), I want to say instead merely that I am in as good a position as anyone to try to keep track of the origins of the things I find myself thinking and doing.

R9. How does neuroscience inform questions of conscious will?

Both Bogen and Raz & Norman complain about the relative lack of neuroscience in ICW. There are no brain scans, and for their money the book’s lack of sophistication in neuroscience is a problem. At one level, this criticism sounds rather egocentric. Just because these folks are neuroscientists, their chosen topic must be covered in detail in any book they read, including one written by a social cognitive psychologist? Do they find it distressing when art museums fail to hang pictures of brains on their walls?

Actually, I think their critique is worth considering – but not because neuroscience should be trotted out these days whenever anyone talks about anything. Rather, it seems that the recent advancement in the study of the brain is involved in motivating the contemporary rethinking of consciousness and conscious will. There is something about seeing our own brains in Technicolor, twinkling with their myriad activations, that prompts a reappraisal of the status of the conscious causer in our heads. The novelist Tom Wolfe (1996) attended a neuroscience conference and summed it up with an essay called Sorry, but your soul just died. Although psychology has long promoted determinism as a way of understanding humans, even offering up radical solutions such as behaviorism, it has never had quite the impact on intuitions about conscious will that is now being driven by pictures of the brain in action.

Indeed, one of the most exciting new ventures in neuroscience is the pursuit of conscious will through scanning methods. It traces the processes whereby conscious will is experienced to cerebellum-mediated internal feedback. A number of other investigators have approached the problem of localizing willed action by examining activations during choice tasks. Brain activity associated with willed action has been explored by comparing self-initiated movement with externally triggered action (Cunnington et al. 2002; Hunter et al. 2003), by comparing perception of one’s own movement and others’ movement (Farrer & Frith 2001), or by comparing normal voluntary movements with movements conducted by patients who have limited experience of voluntariness for the same actions (Spence et al. 2000). These studies often fail to distinguish the experience of conscious will from the functioning of will or choice, and so have not yet yielded an entirely clear picture of how the brain contributes to the feeling of conscious will – as opposed to simply how the brain contributes to choice or action control.

Nonetheless, there is something about the palpable reality of brain activations that makes it a bit easier to envision our own minds as mechanical wonders (rather than magical agents). Perhaps the widespread everyday communication of neuroscience will eventually change our intuitions about conscious will. In a paper examining the impact of neurosciences on moral intuitions, Green and Cohen (in press) wonder exactly this: whether “questions, which seem so important today, will lose their grip in an age when the mechanical nature of decision making is fully appreciated.”

R10. How can we understand responsibility in light of this theory?

How do humans become responsible for their actions if they do not consciously will them? Sternberg puts the question most clearly: “Does Wegner or anyone else want to move to this position – that no one is responsible for his or her own actions?” He decries this theory for departing from common sense, noting that everyone knows that people are responsible for their actions. The related worries of several commentators surface in hand-wringing about not only the moral repercussions of ICW (Heyman), but the religious repercussions, as well (Young, Glymour, Tonnev & Wachholtz). Young wonders alud if this is “a battle that Wegner wishes to fight?” What exactly is at stake?

To begin with, I think we owe nothing to Sternberg’s common sense. Also by someone’s common sense, pictures cannot fly through space, and for that matter, Copernicus had better shut up about the center of the universe. Com-
mon sense is sometimes the enemy, and in this case I think it is gravely misleading.

Qualms about responsibility arise when we make the mistake of believing that responsibility is the same thing as causality. And of course it is not (Hart 1948/1949; 1968). Causality is something you can see in mechanical systems, a relationship between events, and is not dependent on what kinds of events are involved. Responsibility, on the other hand, involves persons — those selves that are constructed through the process of identifying actions as caused by an agent, the “I” (Radden 1996; Wilkes 1988). The creation of a self that can be responsible for anything is a process that is dependent on the feeling of conscious will. Responsibility is created when a person is created. As we have seen, if people say they did not consciously will an action, they have explained that in their view they are not responsible. But this is just their personal estimate of whether the self instantiated in their bodies, brains, and mental processes appeared to them to be causal. And that estimate could be wrong.

The allocation of responsibility is dependent on regarding something as a person, and the first step in this process is that the thing should regard itself as a person. In Dennett’s (Dennett 1987; 1989) terms, this is to say that something is not a person if it does not take the intentional stance toward itself (see also Angel 1989; Scassellati 2002). Because having experiences of conscious will — whether they are correct or incorrect in any given instance — is a necessary step for assigning authorship to oneself, and so doing, fabricating the self, the experience is what creates the possibility of responsibility. Yes, we can hold people or things or events responsible all we want. But this is merely spectator responsibility, the kind of responsibility a judge and jury can give for practical purposes. Only if the accused person accepts the responsibility and experiences moral emotions such as shame or guilt (or in the case of positive actions, pride) will the real work of responsibility be done. We can argue all day about what caused a given action, but this is simply not the same argument as whether the person is responsible.

Responsibility, in this light, is something that exists in the eye of the beholder. Just as consciousness cannot be verified in anyone other than oneself, responsibility is something we each experience. We can then talk about it and make agreements about how much of it a person might have, but these assignments second guess the person’s own experience (Freeman 2000). ICW describes the experience of will as an authorship emotion, a feeling that marks actions whenever they seem to be attributable to self. It is through this process that self is constructed, repeatedly and continually in the course of the day’s actions. Responsibility gives rise to the self, and is not something a self has that gives rise to actions.

How does this view square with Sternberg’s concern that responsibility is being undermined by the theory of apparent mental causation? Indeed, it makes trouble. In the approach of ICW, the whole idea of a “person” is an elegant accounting system for making sense of actions and ascribing them to constructed entities that are useful for purposes of social justice and the facilitation of social interaction. A person is constructed in the mind of the person, and, through a variety of communications and evidences, in the minds of others as well (cf. Dennett 1987). One of the most compelling functions for having conscious will installed in human minds is to give rise to the authorship attribution system we currently have in place, making each person not only understand own authorship but actually feel it whenever actions feel consciously willed. Far from undermining responsibility for action, the mental processes of apparent mental causation function to create such responsibility by making us each feel that we do things. The illusion of conscious will is essential for the development of the first-person sense of responsibility, and this in turn is the basis for the social and legal sense of responsibility we negotiate with each other to achieve effective social relations.

In biology, much theory and research is devoted to questions of how organisms distinguish self from non-self. Complex biological adaptive systems need to be able to make this discrimination at many levels, from the individual molecule to the whole organism, because this discrimination is at the center of successful immune response to antagonists, as well as successful self-organization and development. To date, psychology has not made much of this distinction. Perhaps it is time to start. We should recognize that the experience of conscious will is the start of a self-identification system, a way of tagging actions as belonging to self that, in so doing, creates the self. ICW is not the end of responsibility, but rather a way of modeling how the beginning of responsibility accrues to persons in their own minds.

R11. How should we speak of our selves?

All this talk of selves being destroyed and created is difficult to carry out using normal everyday language. Several commentators who are particularly savvy about this project identify conflicts between what is being said in ICW and the way it is being said, most typically noting that the book talks of “we” or “us” or “I” in ways that promote confusion and occasionally seem ironic. Dennett makes this his central point, asking where the self is in the system that experiences will. He toys with the possible places a self might be, as he has in greater depth elsewhere (Dennett 2003b), and wonders at my saying things like “we inhabit an extraordinarily complicated machine.” Part of this is a personal habit I have of using the indefinite or universal “we,” a location I find particularly useful for talking about psychology (e.g., “We have each had the experience of . . .”). But a larger part of the problem with talking about the self is that it is very difficult to speak of psychological processes without referring to agents or minds by using pronouns.

The shortcomings of language for talking about psychological events and processes are also clear in some of the commentators’ pronouncements. It is all too easy to talk about intrapsychic events sometimes as though they are mechanical events, and other times as though they are the doings or viewings of a person/self. Metzinger says “one right now is a system,” for example, leaping directly from self language to mechanical metaphor in a single phrase. Ainslie talks about functions of will “which depend on our accurate monitoring of them,” and in so doing, runs the equation the other way from system to self. Zuriff says there is “no independent self that makes inferences and experiences itself as causing movements by first having thoughts,” so illustrating the problem again. Velmans expressed this difficulty most directly with his concern that my “first person” analysis cannot be folded back into science. What can be done?

Should everyday language about mental events be jetti-
soned in favor of some kind of self-free rhetoric? Perhaps if psychologists were not allowed to use pronouns in their papers, the problem would go away. No more hocus-pocus, no little people in the head doing or saying or experiencing things! The final curtain could fall on Dennett’s Cartesian Theater. This approach could rapidly remake psychology into a deeply unsatisfying science. One of the great joys of psychological investigation and thought is that we theorists get to trampoline back and forth across the line between subject and object a hundred times every hour, noting in rapid alternation what has caused a person to behave, for example, and what it is like to be the person behaving (Lana 1976; Maslow 1966; Wegner & Gilbert 2000). The wonder of doing science on objects that experience things may simply be too great to give up for the sake of linguistic purity.

William James (1890, p. 138) worried that a desire for precision could end in psychology losing “naturalness of speech,” and asked why we should be asked to forswear the language of our childhood for science. I would hate to start looking over my shoulder and start whispering every time I used a self word to talk about a psychological process. But perhaps that is what it will take (if Dennett is to be satisfied). Ultimately, we may be in for some radically new ways of talking about our selves if we are to understand when it is that selves are actually the topic of conversation, and when we are merely talking about them as a figure of speech. Talking about how our selves come to be may require that we imagine, at least for the sake of argument, that we are not here.

R12. Conclusion

The question period is now at a close, even though many questions less-frequently-asked remain to be broached at all. One last question comes to mind: What have we learned? I for one am impressed with the danger of being self-explanatory. If someone tells you that a software product is self-explanatory, or that a diagram or chart is self-explanatory, that usually means the thing is obvious. Simple. Needs no further ado. But the human mind is self-explanatory as well — in the sense that it presents a simple picture of its own operation to itself. The mind’s model of its operation involves the production of an experience of conscious will for some of its actions and thoughts, and this allows the mind to build a continuing representation of itself as an agent. The self-explanatory property of mind is the basis for the entire matrix of social interaction, the creation of a world of selves who act and interact and ascribe authorship for actions to each other. The self-explanatory mind stands, however, as a significant impediment to psychological scientists bent on discovering how that mind actually works. The illusion of conscious will is both the basis for the construction of persons, and also the rock-hard obvious intuition that often stands in the way of our attempts to understand how persons are constructed.


Carpenter, W. B. (1888) The philosophy of human nature. Appleton. [aDMW]


References/Wegner: Précis of The illusion of conscious will


References/Wegner: Précis of The illusion of conscious will


Q.1. Au: Be the Sparks 1988 reference in your commentary on Wegner. Please check the page range for this reference; currently your reference list provides: 49–5.
This is obviously incorrect. Please send the correct page numbers as soon as possible.

Pylyshyn:
Q. 1. Au: [Hallo again, this is apropos your commentary on Wegner’s 2002 book.] In the Abstract of your commentary on Wegner, re the following sentence with your comment rejecting the previous copyeditor’s edit:
“Examples include the increase in reaction time found when details are reported from smaller images or when attention is switched between images that are further apart.”
I agree that the previous copyeditor’s deletion changes the intended sense and so, in keeping with your request, have rewritten the second part of the sentence, it now reads as follows:
CCE revised version: “Examples include the increase in reaction time found when details are reported from smaller images or when attention is switched between different places and features (imagined as further apart than they are) within a single image.”
PLEASE CONFIRM IF MY REVISION IS OK, RETAINS YOUR INTENDED SENSE? If not, please edit further and send your corrected version back to me asap. [If you’re editing further, make sure the sentence doesn’t get too long, this is the Abstract, needs to be brief.] – S. M. (BBS Chief Copyeditor).

Raz & Norman:
Q.1. Au: In your commentary on Wegner, end of first para. where you mention Wegner’s chapter on hypnosis, please provide the chapter number as well. [Send it by email to Jonathan cc-ing me, Jonathan will insert it in the page proofs.] – S.M. (BBS CCE)
Q.2. Au: Any update by now for the following references cited in your commentary?
Raz (in press)
Raz et al. (in press a)
Raz et al. (in press b)

Schultz et al.:
Q.1. Au: Any update by now for the following references in your commentary on Wegner?

RESPONSE ARTICLE queries:
1. Au: As per previous copyeditor’s note/inserted comment on ms. p. 3 (para beginning: “As a backdrop for these questions . . .”) to the effect whether it is ok to mention commentator John Morton only once, in the Table, and never in the body of the text, my answer is No, it is not ok, please include at least a reference/mention, if not a brief discussion, of John Morton’s commentary or argument within the main body of the Response text. [Sorry, this should have been queried more explicitly, and directly to you, much earlier.] – S.M. (Chief CE)
2. Au: Please check the sense of the following sentence in section R.2:
“If a shooter claims not consciously willing the shot that killed a person, a jury – at least right now, in America – will be at pains to convict the shooter of murder.”
Is something missing? – it seems to me as if the last clause should read: “will be at pains to not convict the shooter of murder” or some tighter version of this, i.e., if a shooter to NOT have consciously willed a killing, wouldn’t a jury normally feel more sympathetic, try to get the shooter off on a lesser charge rather than convict him/her of murder? Your next sentence about manslaughter seems to imply this.
(Or am I for some reason simply getting it all wrong?) – S.M.
3. Au: In section R4 of the Response, paragraph 3, second sentence, where you write:
“From “The feeling of conscious will can be mistaken,” we must conclude that “the feeling of conscious will is never correct.”
Is this a quote from your book, ICW, or from somewhere else? Please provide source or remove quote marks. – S.M.
5. Au: Any update for the Dijksterhuis et al. (2004) reference marked as “manuscript submitted for publication” in your Response reference list? If still in “submitted” stage, then the reference date should read “submitted” in the Response text as well. – S.M.