The New Unconscious
Ran R. Hassin, James S. Uleman, and John A. Bargh

THE NEW UNCONSCIOUS

Edited by
Ran R. Hassin
James S. Uleman
John A. Bargh

OXFORD
UNIVERSITY PRESS

2005

1

Who Is the Controller of Controlled Processes?

Daniel M. Wegner

We are the robots,
We are the robots.
We're functioning automatic,
And we are dancing mechanic.

-"The Robots," Kraftwerk (1978)

Are we the robots? This question surfaces often in current psychological research, as various kinds of robot parts-automatic actions, mental mechanisms, even neural circuits-keep appearing in our explanations of human behavior. Automatic processes seem responsible for a wide range of the things we do, a fact that may leave us feeling, if not fully robotic, at least a bit nonhuman. The complement of the automatic process in contemporary psychology, of course, is the controlled process (Atkinson & Shiffrin, 1968; Bargh, 1984; Posner & Snyder, 1975; Shiffrin & Schnieder, 1977), and it is in theories of controlled processes that vestiges of our humanity reappear. Controlled processes are viewed as conscious, effortful, and intentional. and as drawing on more sources of information than automatic processes. With this power of conscious will, controlled processes seem to bring the civilized quality back to psychological explanation that automatic processes leave out. Yet by reintroducing this touch of humanity, the notion of a controlled process also brings us within glimpsing range of a fatal theoretical error-the idea that there is a controller.

This chapter begins by examining why the notion of a controller is a problem. As we shall see, theories of controlled processes often imply that the person (or some other inner agent such as "'consciousness" or "the will" or "the self") is a legitimate possible cause of the person's observed thought or behavior. This supposition undermines the possibility of a scientific theory of psychology by creating an explanatory entity that cannot itself be explained. The task, then, is to examine how controlled processes might work if they are not the acts of such an agent. What happens if indeed there is no controller? The chapter's conclusion, if you prefer reading from back to front, is this: The operation of controlled mental processes is part of the mental mechanism that gives rise to a sense of conscious will and the agent self in the person. Controlled processes do not start with a controller-in other words, they result in one.

The Homunculus problem

In the film *Manhattan*, Woody Allen searches at one point for the right word to describe his romantic rival. He finally ends up calling him "this little. . . homunculus." He is referring to the rival's stature, in the sense that a homunculus is a little person, a dwarf or manikin. In the sciences of the mind, however, the term is pejorative in quite a different sense. It stands for an absurd explanation-an inner executive agent who "does" the person's actions. Freud's theory of id, ego, and superego, for instance, has often been criticized as a homunculus-based explanatory system in which the person's behavior is explained by reference to an inner agent (in this case, a committee of them) that is responsible for the person's actions. Whenever we explain a behavior by saying that some personlike agent inside the person simply caused it, we have imagined a homunculus and have thereby committed a classic error of psychological explanation.

The issue here, of course, is that a homunculus must itself be explained. The path of explanation implied by the homunculus idea is to reapply the same trick and suggest that another smaller homunculus might be lurking inside the first. This path leads to the specter of an infinite regress of homunculi, nested like Russian dolls, that quickly descends into absurdity. Another way to explain a homunculus is simply to say that it has free will and can determine its own behavior. This means the homunculus causes things merely by deciding, without any prior causes leading to these decisions, and thus renders it an explanatory entity of the first order. Such an explanatory entity may explain lots of things, but nothing explains it. This is the same kind of explanation as saying that God has caused an event. A first-order explanation is a stopper that trumps any other explanation, but that still may not explain anything at all in a predictive sense. Just as we cannot tell what God is going to do, we cannot predict what a free-willing homunculus is likely to do either. There cannot be a science of this,

Most psychologists and philosophers are well aware of the homunculus problem (e.g., Baars, 1997; Dennett, 1978), and it has been generally avoided in contemporary theorizing, with one noteworthy exception. The notions of controlled and automatic processes carry with them the implicit assumption of a kind of homunculus. Now it is true, of course, that most current cognitive and social cognitive research focuses specifically on the automatic side of this dichotomy, so much so that there seems to be progressively less room for the "little person in the head" (e.g., Bargh, 1997; Bargh & Chartrand, 1999; but see Bargh & Ferguson, 2000). But why should there be any room at all? The steady march of automaticity findings is only interesting and understandable because it occurs in the context of its complement, the controlled process. And the controller of controlled processes all too often resembles that inexplicable mini-me, the homunculus.

The homunculus in controlled processing is usually implicit rather than explicit. No theorist has actually said "and therefore, the little person in the head is responsible for the nonautomatic processes we have observed," Baumeister (2000, p. 25) has come very close to this by saying "the self is the controller of controlled processes," and it is not clear what this could mean other than that there is a homunculus to be found controlling things. However, this is a rare expression of what is usually an unspoken assumption, a background belief that controlled processes are somehow more personlike than automatic processes, representing the work of a human agent rather than that of some kind of mechanism. Controlled processes are often seen as conscious, moral, responsible, subtle, wise, reflective, and willful. not because they are described as such in so many words, but rather because they are what is left when we subtract the automatic processes. Everyone knows, after all, that automatic processes may be unconscious, unintentional. primal. and simple-minded-as well as impulsive, selfish, and prejudiced to boot.

The human and the robot inside each person have traditionally been characterized as two different personalities in the person. This was a popular line of theory in early psychology, exemplified by commentators such as Sidis (1898), who described what he called the "subliminal self." This 'unconscious netherself carries out automatic behaviors (such as those suggested by hypnotists) and performs simple-minded actions. Contrasted with this, of course, is the conscious self, capable of all of the fine and intelligent thoughts and actions that any human or homunculus could desire. The subliminal self was robotic and in its simplicity did not need to be viewed as having a mind. In more contemporary discussions, the automatic or unconscious self continues to be appreciated as little smarter than a bar code reader (e.g., Greenwald, 1992), while the conscious self is still accorded full status as a human agent. These comparisons of automatic processes with the processes of the conscious self leave us marveling at the wonders of which the conscious self is capable. They do not, however, explain the conscious self.

Unfortunately, psychology's continued dependence on some version of a conscious self makes it suspect as a science. In the halls of science, controlled processes are haunted by the controller. They seem to have lower status as scientific explanations than automatic processes because there is a ghost in their machine. Automatic processes are seen as more scientifically authentic, reflecting the true nature of humans rather than their conscious and strategic affectations. So, we grimly but readily accept evidence indicating that automatic processes express racial prejudices (e.g., Banaji & Dasgupta, 1998; Devine, 1989) and render careless judgments (Gilbert, 1989). We accept as well that automatic processes prompt blind mimicry of others' behaviors (e.g., Bargh, Chen, & Burrows, 1996; Dijksterhuis & van Knippenberg, 1998), and we readily assent, too, that automatic processes are devilishly difficult to overcome (Macrae, Bodenhausen, Milne, & Jetten, 1994; Wegner, 1994). Automatic processes seem robotic and deeply "causaL" Controlled processes, for their part, seem less than genuine, reflecting unpredictable human choices rather than scientifically respectable causes.

The temptation to imagine a controller seems to be fueled by our deep appreciation of the idea of mind. Early in life, we develop the tendency to understand events that are attributable to minds, and to distinguish them from events that are caused by mechanical processes. The studies of Heider and Simmel (1944), for example, highlighted our extraordinarily compelling inclination to perceive even cartoon geometrical figures as causal agents. The tendency for people to anthropomorphize physical objects and events is a further expression of this natural proclivity (e.g., Guthrie, 1993), and contemporary research on the development of theory of mind in animals and humans suggests that this faculty for mind perception is a strong guiding force in perception more generally (e.g., Carey, 1996).

Our readiness to perceive minds behind events is enhanced further by the experiences we have of our own minds-particularly, the experience of causal agency. We each have extensive experience with the sense that we control our actions, from finger wags to the grandest gestures, and these many instances add up to the convincing intuition that we are controllers who cause our actions. Regardless of whether this feeling that we are doing things is a valid indicator of control. it is this feeling that we tend to equate with the idea of control and that gives us the further intuition that there is always an agent behind the processes that control human thought and action. For controlled processes to reach their full scientific utility, though, they need to be understood apart from any notion of a controller. Controlled processes can indeed be understood as mechanistic processes-for example, as in the cybernetic and dynamical processes posited in control theories (e.g., Bargh & Ferguson, 2000; Carver & Scheier, 1998; Miller, Galanter, & Pribram, 1960; O'Reilly, Braver, & Cohen, 1999; Vallacher & Wegner, 1985; Wegner & Bargh, 1998). These approaches examine the nature of controlled processes

without positing a controller. This is the way it needs to be for progress in the explanation of human psychology. The agent self cannot be a real entity that causes actions, but only a virtual entity, an apparent mental causer.

The controller, in this light, is a personal construction that blends into a scientific illusion. The sense we each have that we are agents who cause our actions is constructed and installed on an act-by-act basis each time we experience causing our action. This experience is available to us primarily during the operation of controlled processes. The personal experience of agency is not a good foundation for a science of mind, however, and we must be careful as scientists to appreciate the basis of this feeling rather than to incorporate the feeling in our theories. To free controlled processes of the controller, it is important to examine how it is that people come to see themselves as controllers. We need to explore the genesis of the experience of conscious will.

Apparent Mental Causation

Why does it feel as though we are doing things? The experience of consciously willing our actions seems to arise primarily when we believe our thoughts have caused our actions. This happens when we have thoughts that occur just before the actions, when these thoughts are consistent with the actions, and when other potential causes of the actions are not present. A theory of apparent mental causation (Wegner, 2002; Wegner & Wheatley, 1999) suggests that these principles of priority, consistency, and exclusivity govern the inferences people make about the causal influence of their thoughts on their actions, and thus underlie the experience of doing things on purpose. In essence, the theory suggests that we experience ourselves as agents who cause our actions when our minds provide us with previews of the actions that turn out to be accurate when we observe the actions that ensue..

Consider some examples to illustrate these principles. If you think of standing up and walking to the window, and then find yourself doing this, the appearance of the thought in mind appropriately prior to the action would support your inference that your thought caused your action. Imagine, however, finding that you had walked to the window without any preview. Having the thought appear only after the action was complete would undermine your experience of will. For that matter, an experience of will would also be subverted if the thought appeared long before the action and then was lost to consciousness by the time you walked to the window. Again, it would not feel as if you had done it, and you might then wonder how you got there. The experience of involuntariness in hypnosis and in automatisms such as Ouija board spelling, water dowsing, and automatic writing can be traced to anomalies of priority (Wegner, 2002). In each case, actions occur without thoughts of them occurring just beforehand, and as a result the actions are experienced as unwilled. The priority principle suggests that the thought must appear in a timely way just before the action for the action to be experienced as voluntary, so departures from this sequence lead to experiences of involuntariness.

In a study of this principle, Wegner and Wheatley (1999) presented people with thoughts (e.g., a tape-recorded mention of the word *swan)* relevant to their action (moving an onscreen. cursor to select a picture of a swan). The movement the participants performed was actually not their own, as they shared the computer mouse with an experimental confederate who gently forced the action without the participants' knowledge. (In yet other trials, the effect of the thought on the participant's own action was found to be nil when the action was not forced.) Nevertheless, when the relevant thought was provided either 1 or 5 seconds before the action, participants reported feeling that they acted intentionally in making the movement. This experience of will followed the priority principle. This was clear because on other trials, thoughts of the swan were prompted 30 seconds before the forced action or 1 second afterward-and these prompts did not yield an inflated experience of will. Even when the thought of the action is wholly externalappearing as in this case over headphones-its timely appearance before the action leads to an enhanced experience of apparent mental causation.

The second key to apparent mental causation is the consistency principle, which describes the semantic connectedness of the thought and the action. Thoughts that are relevant to the action and consistent with it promote a greater experience of mental causation than thoughts that are not relevant or consistent. So, for example, having the thought of eating a salad (and only this thought) just before you find yourself ordering a plate of fries is likely to make the ordering of the fries feel foreign and unwilled (Where did these come from?). Thinking of Ides and then ordering fries, in contrast, will prompt an experience of will. As another example, consider what happens when people with schizophrenia experience hearing voices. Although there is good evidence that these voices are self-produced, the typical response to such auditory hallucinations is to report that the voice belongs to someone else. Holman (1986) has suggested that the inconsistency of the utterance with the person's prior thoughts leads to the inference that the utterance was not consciously willed-and so to the delusion that others' voices are speaking "in one's head." Ordinarily, we know our actions in advance of their performance and experience the authorship of action because of the consistency of this preview with the action.

In a laboratory test of the consistency principle, Wegner, Sparrow, and Winerman (2004) arranged for each of several undergraduate participants to observe their mirror reflection as another person behind them, hidden from view, extended arms forward on each side of them. The person behind the participant then followed instructions delivered over headphones for a series of hand movements. This circumstance reproduced a standard pantomime sometimes called Helping Hands in which the other person's hands look, at least in the mirror, as though they belong to the participant. This appearance did not lead participants to feel that they were controlling the hands if they only saw the hand movements. When participants could hear the instructions that the hand helper followed as the movements were occurring, though, they reported an enhanced feeling that they could control the other's hands.

In another experiment on hand control. this effect was again found. In addition, the experience of willing the other's movements was found to be accompanied by an empathic sensation of the other's hands. Participants for this second study watched as one of the hands snapped a rubber band on the wrist of the other, once before the sequence of hand movements ..and once again afterward. All participants showed a skin conductance response (SCR) to the first snap-a' surge in hand sweating that lasted for several seconds after the snap. The participants who had heard previews of the hand movements consistent with the hands' actions showed a sizeable SCR to the second rubber band snap as well. In contrast, those with no previews, or who heard previews that were inconsistent with the action, showed a reduced SCR to the snap that was made after the movements. The experience of controlling the hand movements seems to induce a sort of emotional ownership of the hands. Although SCR dissipated after the movements in participants who did not hear previews, it was sustained in the consistent preview condition. The consistency of thought with action, in sum, can create a sense that one is controlling someone else's hands and, furthermore, can yield a physiological entrainment that responds to apparent sensations in those hands. It makes sense in this light that consistency between thought and action might be a powerful source of the experience of conscious will we feel for our own actions as well.

The third principle of apparent mental causation is exclusivity, the perception that the link between one's thought and action is free of other potential causes of the action. This principle explains why one feels little voluntariness for an action that was apparently caused by someone else. Perceptions of outside agency can undermine the experience of will in a variety of circumstances, but the most common case is obedience to the instructions given by another. Milgram (1974) suggested in this regard that the experience of obedience introduces "agentic shift"-a feeling that agency has been transferred away from oneself. More exotic instances of this effect occur in trance channeling, spirit possession, and glossolalia or "speaking in tongues," when an imagined agent (such as a spirit, entity, or even the Holy Spirit) is understood to be influencing one's actions, and so produces a decrement in the experience of conscious will (Wegner, 2002).

A further example of the operation of exclusivity is the phenomenon of facilitated communication (FC), which was introduced as a manual technique for helping autistic and other communication-impaired individuals to communicate without speaking, A facilitator would hold the client's finger above a letter board or keyboard, ostensibly to brace and support the client's pointing or key-pressing movements, but not to produce them. Clients who had never spoken in their lives were sometimes found to produce lengthy typed expressions this way, at a level of detail and grammatical precision that was miraculous. Studies of FC soon discovered, however, that when separate questions were addressed (over headphones) to the facilitator and the client, those heard only by the facilitator were the ones being answered. Facilitators commonly expressed no sense at all that they were producing the communications, and instead they attributed the messages to their clients. Their strong belief that FC would work, along with the conviction that the client was indeed a competent agent whose communications merely needed to be facilitated, led to a breakdown in their experience of conscious will for their own actions (Twachtman-Cullen, 1997; Wegner, Fuller, & Sparrow, 2003). Without a perception that one's own thought is the exclusive cause of one's action, it is possible to lose authorship entirely and attribute it even to an unlikely outside agent.

Another example of the exclusivity principle at work is provided in studies of the subliminal priming of agents (Dijksterhuis, Preston, Wegner, & Aarts, 2(04). Participants in these experiments were asked to react to letter strings on a computer screen by judging them to be words or not-and to do this as quickly as possible in a race with the computer. On each trial in this lexical decision task, the screen showing the letters went blank either when the per- . son pressed the response button, or automatically at a short interval (about 400-650 ms) after the presentation. This made it unclear whether the person had answered correctly and turned off the display or whether the computer did it, and on each trial the person was asked to guess who did it. In addition, however, and without participants' prior knowledge, the word I or me or some other word was very briefly presented on each trial. This presentation lasted only 17 ms, and was both preceded and followed by random letter masks-such that participants reported no awareness of these presentations.

The subliminal presentations influenced judgments of authorship. On trials with the subliminal priming of a first-person singular pronoun, participants more often judged that they had beaten the computer. They were influenced by the unconscious priming of self to attribute an ambiguous action to their own will. In a related study, participants were subliminally primed on some trials with the thought of an agent that was not the self-God. Among those participants who professed a personal belief in God, this prime reduced the causal attribution of the action to self. Apparently, the decision of whether self is the cause of an action is heavily influenced by the unconscious accessibility of self versus nonself agents. This suggests that the exclusivity of conscious thought as a cause of action can be influenced even by the unconscious accessibility of possible agents outside the self.

The theory of apparent mental causation, in sum, rests on the notion that our experience of conscious will is normally a construction. When the right timing, content, and context link our thought and our action, this construction yields a feeling of authorship of the action. It seems that we did it. However, this feeling is an inference we draw from the juxtaposition of our thought and action, not a direct perception of causal agency. Thus, the feeling can be wrong. Although the experience of will can become the basis of our guilt and our pride, and can signal to us whether we feel responsible for action in the moral sense as well, it is merely an estimate of the causal influence of our thoughts on our actions, not a direct readout of such influence. Apparent mental causation nevertheless is the basis of our feeling that we are controllers.

From Controlled Processes to an Agent

The feeling of conscious will that occurs with any given action is likely to be influenced by the psychological process responsible for that action. If the process allows access to information indicating that thoughts occurred with appropriate levels of priority, consistency, and exclusivity, the action will be experienced as willed, whereas in other cases it will not. So, for instance, psychological processes that create snoring when we are asleep might yield a particularly impoverished array of information and/or computational ability regarding will-and so fail to create an experience that we are snoring on purpose. The information in this case fails to establish conscious will because, in sleep, we do not even have conscious thoughts that can be assessed for their relative priority, consistency, and exclusivity, nor are we likely to be doing much computation.

This observation suggests that variability in the experience of conscious will may be attributable to variations in the availability of the essential sources of information for the computation of apparent mental causation, as well as the availability of mental resources. Such availability could flow from the very processes creating action. Certain processes, then, allow the experience of will, while others do not. In particular, the operation of will may be inferred to the degree that there are available (1) conscious thoughts, (2) observable actions, and (3) time and attention to infer a causal link between them. In each of these respects-conscious thinking, action monitoring, and attention deployment-controlled processes are more likely to support an inference of conscious will than are automatic processes. This line of reasoning suggests how it is that controlled processes can create the experience of a controller.

Consider first the role of conscious thinking. If a person has no conscious thoughts prior to an action, apparent mental causation cannot be inferred. The idea that controlled processes are conscious maps onto this criterion directly. Indeed, it is difficult to imagine a controlled process that ensues without some kind .of conscious preview, so much so that it is common to find the term *conscious* substituting for *controlled* and compared with *automatic* in the research literature (e.g., Bargh, 1994; Wegner & Bargh, 1998). If controlled processes involve conscious thought, while automatic processes need not do so, then it is primarily through controlled processes that a controller might be inferred. The same reasoning 'applies when we consider the intentionality of the controlled process. In emphasizing that controlled processes are intentional, Bargh (1994) and most other commentators allow that the intention may cause the action (and by this suggestion, they breathe a bit of life into the controller). Bargh and Ferguson (2000) and Wegner and Bargh (1998) have recognized this problem and have suggested that concepts of control and intention need to be defined without reference to a controlling agent. The apparent mental causation perspective that follows from this view suggests that the intention is important not as a cause, but because it is a conscious preview of the action that is often consistent with, the action that is subsequently observed. The intent that precedes or accompanies a controlled process thus serves as a basis for an experience of will and enables the inference of a controller.

The second feature of controlled processes that allows us to infer an agent is the fact that they involve action monitoring. Automatic processes, of course, are generally understood as unmonitored or even ballistic-processes that once started cannot be stopped or even guided. The outputs of automatic processes may thus never be known to the person. Controlled processes, in contrast, typically involve a feedback loop, a comparison between what was intended and what actually happened. This comparison requires that the person know at some level, sometimes consciously, what action. has indeed occurred. And it is this conscious monitoring of the completed action that is necessary for the inference of apparent mental causation. One can only feel will for actions one knows one has performed. This means that many of the automatic actions observed in psychological laboratories cannot give rise to any inference of mental causation. Unless participants in a study specifically know that they are walking more slowly, for example (see Bargh et aI., 1996), they will not be able to infer that they consciously willed doing this. When automatic processes do happen to announce their resulting actions and thoughts to consciousness, they may then be eligible to give rise to a sense of agency. But controlled processes do this every time.

The third feature of controlled processes that supports the inference of agency is the degree of attention deployment they allow. Controlled processes are typically marked by slowness and thoroughness, as the attention devoted to them makes them both resource draining and methodical. Consider their slowness first. Time to think is particularly useful when causal inferences need to be made regarding one's own thought and action, and automatic processes are not likely to provide this time. Automatic processes can yield results, often in milliseconds, whereas controlled processes may take days and at a minimum seem to require several hundred milliseconds. Responding to a green light by punching the accelerator, for instance, can occur almost before we' are conscious that the light is green. In a study in which participants were tracking by hand an unexpectedly moving target, for example, the change in their hand trajectory toward the target's movement happened as early as 100 ms following the target jump. However, the voc<\l signal by which they reported their consciousness of the jump (in this case, saying "Pah") did not occur on average until more than 300 ms later (Castiello, Paulignan, & Jeannerod, 1991). The sheer speed of automatic processes leaves inferences of agency in the dust.

The thoroughness of controlled processes is related to their use of attention deployment as well. A conscious half hour meandering through all the possible responses one might make to an insult, for example, is likely *to* produce a far more thorough, studied, and balanced response than is a quick, automatic retort. The thoroughness of controlled processes allows them to review and integrate a far wider range of information on the way to their output than is the case for automatic processes. It makes sense, then, that the reasoning involved in examining priori(y, consistency, and exclusivity for an action is more likely to be developed through controlled processes than through automatic processes.

Because automatic actions do not support inferences of agency during the action, it turns out that many of our most fluid, expert, and admirable actions are ones we do not experience consciously willing. Should we write a particularly beautiful piece of prose, there is often a distinct sense that it happened to us rather than that we did it. Scientists and mathematicians similarly claim that their creative discoveries seemed just to pop into their heads (e,g., Koestler, 1989). This loss of the sense of authorship in skilled action occurs widely in sports as well, such that calling a player "unconscious" turns out to be a major compliment. Admittedly, most people are quite willing to take credit for skilled actions after the fact, as few writers, scientists, or sports stars turn down their paychecks. But the intriguing aspect of automatic actions is that' they do not feel willed as they unfold. Because appropriate previews do not seem to come to mind to allow the inference of conscious will, the authors of skilled actions often report feeling like spectators who happen to have particularly good seats to view the action.

Feelings of conscious will are most likely, in this view, for actions that we traditionally understand as involving "willpower." When our thoughts about an action appear very prominently before the action occurs-such as when we ponder in depth our plan to resist that drink or smoke or extra muffin we then experience an unusual surge of the feeling of will. The exercise of self-control creates an apt circumstance for this feeling because it specifically involves an intense preview period just before the action (of resistance). When we succumb to some automatic or habitual indulgence, in contrast, we seldom think much about it, and so we experience little sense of willing the indulgent act. Automatic indulgences tend to occur when we have thought of the action long in advance of its occurrence (such as when we premeditate dropping by the bar on the way home just in case some friends are drinking), or when we think of the action as it occurs (such as when we're putting the drink to our lips). The optimal time for thought that contributes to feelings of will is a few moments before the action, and this is when our thoughts of moderation, when effective, can yield great waves of will and resultant self congratulation.

These observations suggest that we feel conscious will as we perform our actions primarily in the case of actions that are caused by controlled processes. These processes allow us the conscious thoughts, self-observed actions, and time and attention necessary to draw causal inferences about how our minds seem to be involved in producing our behaviors. In drawing these inferences, we accumulate the picture of a virtual agent, a mind that is apparently guiding the action. Although this mind is a deeply important construction, allowing us to understand, organize, and remember the variety of things we find ourselves doing, it is a construction nonetheless and must be understood as an experience of agency derived from the perception of thoughts and actions-not as a direct perception of an agent.

Virtual Agency

The creation of our sense of agency is critically important for a variety of personal and social processes, even if this perceived agent is not a. cause of action. The experience of conscious will is fundamentally important because it provides a marker of our authorship-what might be called an authorship emotion. In the words of T. H. Huxley (1910, p. 218), "Volition. . . is an emotion indicative of physical changes, not a cause of such changes." Each surge of will we sense in the operation of controlled processes provides a bodily reminder of what we think we have done. In this sense, the function of will is to identify our actions with a feeling, allowing us to sense in a very basic way what we are likely to have done, and to distinguish such things from those caused by events in the world or by other people or agents. Like the somatic marker function of emotion (Damasio, 1994), the experience of conscious will anchors our actions to us in a way that transcends rational thought.

Conscious will is a cognitive feeling, like confusion or the feeling of knowing (see Clore, 1992). Although it does not have an associated facial expression, it shares with the basic emotions an experiential component-we do not just deduce that we did an action, we feel that we did it. 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, remembering them and organizing them into a coherent picture of our own identity as agents. By this reasoning, conscious will can be understood as part of an intuitive accounting system that allows us to deserve things. We must know what we have done if we are going to claim that our actions have earned us anything (or have prevented us from deserving something nasty). Our sense of what we have achieved, and our ideas, too, of what we are responsible for in moral domains, may arise because we gain a deep apprehension of our likely causal role in the experience of will.

The creation of personal action authorship must thus be attributable to controlled processes. This means that automatic processes regularly fail to create an agent self, the sense that there is an "{" who did the action. Automatic processes seem to emanate from an unperceived center, a seemingly robotic source that does not experience its own likely complicity in action causation. Automatic processes can occur and leave us like zombies, often not knowing our actions in advance or afterward, and also without the mental resources to compute our complicity. Automatic processes leave us in the dark that we are authors at all. In promoting an experience of will, in turn, controlled processes allow us to experience the subjective causation of the controlled action, and so open the door to the experience of personal emotions such as pride and disappointment in achievement domains, not to mention moral emotions such as guilt and elevation in moral domains (cf. Uleman, 1989).

Because controlled processes give rise to the sense of authorship, they open up the possibility that thoughts of authorship can influence subsequent action. Controlled processes leave a residue of memories of past authorship, and give rise as well to anticipations of future authorship. It is in this sense that many theorists have spoken of the self as a kind of narrator, creating a life story (e.g., Dennett, 1992). Controlled processes can take authorship issues into account because authorship has been created by other controlled processes in the past and can be anticipated to arise from controlled processes in the future. The experiences of authorship that enter into action this way are not direct perceptions of an agent, it should be remembered, but rather are estimates of the role of one's own thoughts in action that are produced by the system that infers apparent mental causation. Far from a simple ho humunculus that "does things," then, the self can be understood as a system that arises from the experience of authorship, and is developed over time by a set of controlled processes that manage memories and anticipations of authorship experiences. We become agents by experiencing what we do, and this experience then informs the processes that determine what we will do next.

Yes, this line of thinking does seem a bit cumbersome when we compare it to the naive simplicity of homunculus talk. To be accurate, we must speak of apparent mental causation, or of virtual agency, rather than of intention or of a controller. But the labor we expend to keep the controller out of our theorizing may well repay us with new insights into phenomena that were impenetrable given only the notion that there is a little person in the head. Profound mysteries in the psychology of self and identity might become a bit less mysterious. We might begin to appreciate, for example, how there could ever be multiple little people (as in multiple personality disorder) or how there could be replacement little people (as in spirit possession or channeling) or even how each person's own sense of the little person inside (as in the development of the agent self) might become open to more effective explanation. On the fringes of our current understanding lie many phenomena that have not been tractable because the assumption of a real controller makes them seem quite out of the question. These phenomena do seem to exist, and our further thinking about the nature of control, automaticity, and the self can be informed by them. All we need to do is assume, for sake of argument, that we are the robots.

Real Mental Causation

In focusing on the topic of apparent mental causation, this chapter has tiptoed quietly around the big sleeping problem of real mental causation. Questions of whether thought actually does cause action, for example, have been left in peace, and the issue of the role of consciousness in the causation of action has been ignored as well. This is because the focus of this theory is the experience of conscious will, not the operation of the will. According to this theory, the experience of 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, then, and not in accord with any actual causal link between thought and action. This theory is mute on whether thought does cause action.

Most theories of behavior causation have gotten this all confused. Questions of how thought or consciousness might cause. action have been muddled together with questions of the person's experience of such causation, and in this snarl nothing seems particularly clear. In large part, this seems to have happened because the feeling of free will is so deeply powerful and impressive. All too often, we take as gospel truth our personal intuition that our conscious thoughts cause our actions ("See, I'm moving my finger!"), and we assume that this experience is a direct pipeline to the truth of the matter.

But imagine for a minute that we are robots. Imagine that our actions arise from a complicated set of mechanisms. Imagine, too, that these excellent mechanisms also give rise to thoughts about what we will do that preview our actions quite reliably. In other words, all the trappings are present to allow us to experience apparent mental causation. If we were robots, would our reports of willed actions (e.g., "I raised my finger") be understood as infallible indicators of the actual causal sequence underlying our actions? If someone had installed a will-interpretation mechanism that used our thoughts of actions and our actions to infer our authorship (e.g., "Thought of finger raising occurred 800 ms before finger went up"), would the output of that mechanism be considered a direct readout of how the action had been produced?

Far from it. The robot analysis team down at the factory would take this output as one piece of evidence, but would want to have tests and sensors and gauges in every sprocket to discern whether this potential causal path was indeed the right one. In this sense, reports of apparent mental causation from humans and robots alike should be taken as estimates of the underlying mechanism at best-and certainly not as readouts of the causal mechanism underlying actions. The way the mind seems to its owner is the owner's best guess at its method of operation, not a revealed truth.

Notes

[1] The "self" in Baumeister's theory is more a repository of prior causal influences than it is a homunculus or uncaused cause. So I am probably picking on him unfairly in singling out this statement. Still, such talk of a controller certainly prompts images of a little person.

[2] A note on definition is helpful here. Attempts to define automatic and controlled processes have pointed to several features that seem to distinguish them. Bargh (1994) suggested that the processes tend to differ in their susceptibility to consciousness, ability to be intended, efficiency, and susceptibility to inhibition. Automatic processes do not usually have all of these features-they are not simultaneously unconscious, unintended, efficient, and unstoppable-and instead seem to be defined as having at least one of the features. Controlled processes. on the other hand, regularly do seem to share all the complementary features-they are conscious, intended, inefficient, and stoppable. Wegner and Bargh (1998) suggested this asymmetry reveals that controlled processes are the defming end of this dimension, from which automatic processes are noted for their departure.

References

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2, pp. 89-195). New York: Academic Press.

Baars, B. J. (1997). *In the theater of consciousness.* New York: Oxford University Press.

Banaji, M. R.. & Dasgupta, N. (1998). The consciousness of social beliefs: A program of research on stereotyping and prejudice. In V. Y. Yzerbyt, B. Dardenne, & G. Lories (Eds.), *Metacognition: Cognitive and social dimensions* (pp.157-170). Thousand Oaks, CA: Sage.

Bargh, J. A. (1984). Automatic and conscious processing of social information. In R. S. Wyer Jr. & T. K. Srull (Eds.), *Handbook of social cognition* (Vol. 3, pp. 1-43). Hillsdale, NJ: Erlbaum. .

Bargh, J. A. (1994). The four horsemen of automaticity: Awareness, intention, efficiency, and control. In R. S. Wyer Jr. & T. K. Srull (Eds.), *Handbook of social cognition* (2nd ed., Vol. 1. pp. 1-40). Hillsdale, NJ: Erlbaum.

Bargh, J. A. (1997). The automaticity of everyday life. In R. S. Wyer, Jr. (Ed.), *The automaticity of everyday life: Advances* in *social cognition* (Vol. 10, pp. 1-61). Mahway, NJ: Erlbaum.

Bargh, J. A., & Chartrand, T. 1. (1999). The unbearable automaticity of being. *American Psychologist,* 54, 462-479.

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology,* 71,230-244.

Bargh, J. A., & Ferguson, M. J. (2000). Beyond behaviorism: On the automaticity of higher mental processes. *Psychological Bulletin,* 126, 925-945. Baumeister, R. F. (2000). Ego depletion and the self's executive function. In A. Tesser, R. B. Fe]son, & J. M. Suls (Eds.), *Psychological perspectives on self and identity* (pp. 9-33). Washington, DC: American Psychological Association.

Carey, S. (1996). Cognitive domains as modes of thought. In D. R. Olson & N. Torrance (Eds.), *Modes of thought: Explorations in culture and cognition* (pp. 187-215). New York: Cambridge University Press.

Carver, C. S., & Scheier, M. F. (1998). On *the sefregulation of behavior.* New York: Cambridge University Press.

Castiello, U., Paulignan, Y., & Jeannerod, M..(1991). Temporal dissociation of motor responses and subjective awareness: A study in normal subjects. *Brain,* 114, 2639-2655.

Clore, G. (1992). Cognitive phenomenology: Feelings and the construction of judgment. In L, L. Martin (Ed.), *The construction of social judgments* (pp. 133-163). Hillsda]e, NJ: Erlbaum.

Damasio, A. R. (1994). *Descartes' error: Emotion, reason, and the human brain.* New York: Avon.

Dennett, D. C. (1978). Toward a cognitive theory of consciousness. In D. C. Dennett (Ed.), *Brainstorms* (pp. 149-173). Cambridge, MA: Bradford Books/MIT Press.

Dennett, D. C. (1992). The self as a center of narrative gravity. In F. Kessel. P. Cole, & D. Johnson (Eds.), *Self and consciousness: Multiple perspectives* (pp.l03-115). Hillsdale, NJ: Erlbaum.

Devine, P. G, (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology,* 56, 680-690.

Dijksterhuis, A., & van Knippenberg, A. (1998). The relation between perception and behavior or how to win a game of Trivial Pursuit. *Journal of Personality and Social Psychology,* 74, 865-877.

Dijksterhuis, A., Preston, J.. Wegner, D. M., & Aarts, H. (2004). *Effects of subliminal priming of natural and supernatural agents on judgments of authorship.* Manuscript submitted for publication.

Gilbert, D. T. (1989). Thinking lightly about others: Automatic components of the social inference process. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 189-211). New York: Guilford.

Greenwald, A. G. (1992). New Look 3: Unconscious cognition reclaimed. *American Psychologist,* 47, 766-779.

Guthrie, S. E. (1993). *Faces* in *the clouds: A new theory of religion.* New York: Oxford University Press,

Heider, F.. & Simmel. M. (1944). An experimental study of apparent behavior. *American Journal of Psychology,* 57, 243-259,

Hoffman, R. E. (1986). Verbal hallucinations and language production processes in schizophrenia. *Behavioral and Brain Sciences,* 9, 503-548.

Huxley, T. H. (1910). *Methods and results.* New York: Appleton *Co.*

Koestler, A. (1989). *The act of creation.* London: Arkana/Penguin.

Macrae, C. N., Bodenhausen, G, V., Milne, A. B., & Jetten, J. (1994). Out of mind but back in sight: Stereotypes on the rebound. *Journal of Personality and Social Psychology,* 67, 808-817.

Milgram, S. (1974). *Obedience to authority.* New York: Harper and *Row.*

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior.* New York: Holt, Rinehart, and Winston.

O'Reilly, R. *C.*, Braver, T. S., & Cohen, J. D.(1999). A biologically-based computational model of working memory. In A. Miyake & P. Shah (Eds.). *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 102-134). New York: Cambridge University Press.

Posner, M. 1., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. So]so (Ed,), *Information processing and cognition* (pp, 55-85), Hillsdale, NJ: Erlbaum.

Shiffrin, R, M., & Schneider, W. (1977). Controlled and automatic human information processing II. Perceptual learning, automatic attending, and a general theory. *Psychological Review,* 84, 127-190.

Sidis, B. (1898). *The psychology of suggestion.* New York: D. Appleton and *Co.* Twachtman-Cullen, D. (1997). *A passion to believe: Autism and the facilitated communication phenomenon,* Boulder, *CO:* Westview.

Uleman, J. S. (1989). A framework for thinking intentionally about unintended thought. In J. S. Uleman & J. A. Bargh (Eds,), *Unintended thought* (pp. 425-449). New York: Guilford.

Vallacher, R. R., & Wegner, D. M. (1985). *A theory of action identification.* Hillsdale, NJ: Erlbaum.

Wegner, D. M. (1994). Ironic processes of mental control. *Psychological Review, 101,* 34-52.

Wegner, D. M, (2002), *The illusion of conscious will,* Cambridge, MA: MIT Press.

Wegner, D, M., & Bargh, J, A. (1998). Control and automaticity in social life. In D, Gilbert, S. T. Fiske, & G, Lindzey (Eds.), *Handbook of social psychology* (4th ed., pp. 446-496). Boston: McGraw-Hill.

Wegner, D. M., Fuller, V, A., & Sparrow, B. (2003). Clever hands: Uncontrolled intelligence in facilitated communication. *Journal of Personality and Social Psychology*, 85, 5-19.

Wegner, D. M., Sparrow, B., & Winerman, L. (2004). Vicarious agency: Experiencing control over the movements of others. *Journal of Personality and Social Psychology*, 86, 838-848.

Wegner, D. M.. & Wheatley, T. (1999). Apparent mental causation: Sources of the experience of will. *American Psychologist*, 54, 480-491.