

# 1

---

## The Illusion

*It usually seems that we consciously will our voluntary actions, but this is an illusion.*

All theory is against the freedom of the will; all experience is for it.

Samuel Johnson,  
Boswell's *Life of Johnson*  
(1791)

So, here you are reading a book on conscious will. How could this have happened? One way to explain it would be to examine the causes of your behavior. A team of scientific psychologists 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 picked up this book at this moment.<sup>1</sup> However, another way to explain the fact of your reading this book is

1. This assumption is similar to a conjecture of the French astronomer and mathematician Pierre Simon Laplace (1749–1827) in his *Philosophical Essay on Probabilities* (1814): “An intellect which at any given moment knew all the forces that animate Nature and the mutual positions of the beings that comprise it, if this intellect were vast enough to submit its data to analysis, could condense into a single formula the movement of the greatest bodies of the universe and that of the lightest atom: for such an intellect nothing could be uncertain; and the future just like the past would be present before its eyes.” It turns out that this “single formula” is so complex that the project of understanding the causation of even a single human action is a vast challenge to scientists, perhaps an impossible one. However, we’re talking here about an ideal of science, not a practical project.

just to say that you decided to pick up the book and begin reading. You consciously willed what you are doing.

These two explanations are both appealing but in different ways. The scientific explanation accounts for behavior as a mechanism and appeals to that part of us that knows how useful science is for understanding the world. It would be wonderful if we could understand people in just the same way. The conscious will explanation, on the other hand, has a much deeper grip on our intuition. We each have a profound sense that we consciously will much of what we do, and we experience ourselves willing our actions many times a day. As William James put it, “The whole sting and excitement of our voluntary life . . . depends on our sense that in it things are *really being decided* from one moment to another, and that it is not the dull rattling off of a chain that was forged innumerable ages ago” (1890, 453). Quite apart from any resentment we might feel on being cast into the role of mechanisms or robots, we appreciate the notion of conscious will because we experience it so very acutely. We do things, and when we do them, we experience the action in such a way that it seems to flow seamlessly from our consciousness. We feel that *we cause ourselves to behave*.

The idea of conscious will and the idea of psychological mechanisms have an oil and water relationship, having never been properly reconciled. One way to put them together—the way this book explores—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. We should be able to examine and understand what creates the experience of will and what makes it go away. This means, though, that conscious will is an illusion.<sup>2</sup> It is an illusion in the sense that *the experience of consciously willing an action is not a direct indication that the conscious thought has caused the action*. Conscious will, viewed this way, may be an extraordinary illusion indeed—the equivalent of a magician’s producing an elephant from the

2. Calling this an illusion may be a bit strong, and it might be more appropriate to think of this as a construction or fabrication. But the term *illusion* does convey the possibility that we place an erroneously large emphasis on how will appears to us and assume that this appearance is a deep insight.

folds of his handkerchief. How could it seem so much like our wills cause our actions if this isn't actually happening? To grasp how this might be, we need to begin by examining what exactly is meant by conscious will. With any luck, we will discover a large expanse of the elephant protruding from the magician's pocket and so begin to understand how the trick works.

### Conscious Will

Conscious will is usually understood in one of two major 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 that this book is about. So, to begin, we'll need to look into each in turn, first examining will as an experience and then considering will as a causal force.

#### *The Experience of Conscious Will*

Will is a feeling. David Hume was sufficiently impressed by this idea so that he proposed to define the will in this way, as “nothing but *the internal impression we feel and are conscious of, when we knowingly give rise to any new motion of our body, or new perception of our mind*” (1739, 399). 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 they've done something that they 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

isn't 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 fifteen minutes in there scrubbing up nicely, and then comes out reporting that she indeed seems to have had a shower but does not feel she had consciously willed it—who are we to say that she did will it? Consciously willing an action requires a feeling of doing (Ansfield and Wegner 1996), a kind of internal “oomph” that somehow certifies authentically that one has done the action. If she didn't get that feeling about her showering, then there's no way we could establish for sure whether she consciously willed it.

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 doesn't 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 instance, 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. One such person was the character played by Peter Sellers in *Dr. Strangelove* (fig. 1.1), who couldn't control one hand and found it alternately steering his wheelchair astray and gesturing a Nazi salute.

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 lobe on the side of the brain opposite the affected hand (Gasquoin 1993), and in some people the difficulty can come and go over time (Leiguarda et al. 1993). Banks and colleagues (1989, 456) 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 grasp 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.”

Should the alien hand's movements be classed as willed or unwilled? On the one hand (pun couldn't be helped), the alien hand seems to do some



**Figure 1.1**  
Dr. Strangelove. Courtesy Archive Photos.

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 right hand while the left one unzipped his jacket; he tried to soap a washcloth while the left hand kept putting the soap back in the dish; and he tried to open a closet with the right hand while the left one closed it" (Banks et al. 1989, 457). 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. One patient described the experience as a feeling that “someone from the moon” was controlling her hand (Geschwind et al. 1995, 803).

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, Rhue, and Weekes 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. You will find 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 down toward the ground.” With enough of this patter, many listeners will indeed experience the arm’s 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 doesn’t 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 and Lynn 1998; Spanos 1986). Hypnotic involuntariness thus provides an example of the lack of experience of will that is even more perplexing than alien hand syndrome. With alien hand, the person simply doesn’t 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, even this foreknowledge of the action seems insufficient to move the action into the “consciously willed” category. If it doesn’t feel as though you did it, then it doesn’t seem that the will was operating.



**Figure 1.2**  
A table-turning séance. From *L'Illustration* (1853).

Another case of the absence of experience of will occurs in *table turning*, a curious phenomenon discovered in the spiritualist movement in Europe and America in the mid-nineteenth century (Ansfield and Wegner 1996; Carpenter 1888; Pearsall 1972). To create this effect, a group of people sit gathered around a table, all 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 (fig. 1.2). It might even move about the room or begin rotating so quickly that the participants can barely keep up. Carpenter (1888, 292–293) observed that “all this is done, not merely without 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.”

In one exemplary case, the Reverend N. S. Godfrey, his wife, and a friend one evening in June 1852 placed their hands on a small mahogany table and found that after forty-five minutes it began to move. With two family servants and the local schoolmaster as witnesses, the group carried out experiments and found that the table would move in various ways, some of which seemed particularly sinister. At one point something

“caused the table to revolve rapidly,” yet then, as Godfrey relates, “a bible was quietly laid upon the table and it stopped! We were horror struck!” (1853, 23). Questions were asked of the table, and responses were given by a leg’s rising and knocking on the floor, and interchanges ensued that convinced those assembled that there was a devil inhabiting the table and causing it to move.

The table-turning curiosity was sufficiently celebrated and controversial to attract 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). All one needs to do, actually, is to use a dusty table and observe the direction of the streaks left by participants’ slipping hands. The streaks run away from their hands in the direction opposite the table movement (as one would expect if people’s fingers slipped a bit as they pushed the table) rather than toward the movement (as one would expect if the table were pulling them along and their fingers were slipping as they fell behind). Apparently, in attributing the table movement to the spirit, the participants did not have sufficient experience of will to recognize the source of their own voluntary actions. Indeed, the Reverend Godfrey disputed Faraday’s findings vehemently: “[We] imparted the motion, he tells us, *which we did not.*”

Such examples of the separation of action from the experience of will suggest that it is useful to draw a distinction between them. Figure 1.3

	Feeling of Doing	No Feeling of Doing
Doing	Normal voluntary action	Automatism
Not Doing	Illusion of control	Normal inaction

Figure 1.3  
Conditions of human action.



shows what might be considered 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 quadrant shows 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 non-controversial case, or perhaps the assumed human condition. The lower right quadrant is also noncontroversial, the instance when we are not doing anything and feel we are not.

The upper right quadrant—the case of no feeling of will when there is in fact action—encompasses the examples we have looked at so far. The movement of alien hands, the hypnotic suggestion of arm heaviness, and table turning all fit in this quadrant because they involve no feeling of doing in what appear otherwise to be voluntary actions. These can be classed in general as *automatisms*.<sup>3</sup> More automatisms are explored in later chapters. The forms they take and the roles they play in life are something of a subtext throughout this book. We should not fail to notice here, however, the other special quadrant of the table—cases of the *illusion of control*. Ellen Langer (1975) used this term to describe instances in which people have the feeling they are doing something when they actually are not doing anything.<sup>4</sup>

When does this happen? The last time it happened to me was when I was shopping in a toy store with my family one Saturday. While my kids were taking a complete inventory of the stock, I eased up to a video game display and started fiddling with the joystick. A little monkey on the screen was eagerly hopping over barrels as they rolled toward him, and I got quite involved in moving him along and making him hop, until the

3. An automatism is not the same thing as an automatic behavior, though the terms arise from the same beginnings (first mentioned by Hartley 1749). An automatism has been defined as an apparently voluntary behavior that one does not sense as voluntary (Carpenter 1888; Janet 1889; Solomons and Stein 1896), and we retain that usage here. More generally, though, a behavior might be automatic in other senses—it could be uncontrollable, unintended, efficient, or performed without awareness, for instance (Bargh 1994; Wegner and Bargh 1998).

4. This term is not entirely fitting in our analysis because the larger point to be made here is that all of the feeling of doing is an illusion. Strictly speaking, then, the whole left side of the table should be labeled illusory. But for our purposes, it is worth noting that the illusion is particularly trenchant when there is intention and the feeling of doing but in fact no action at all.

phrase “Start Game” popped into view. I was under the distinct impression that I had started some time ago, but in fact I had been “playing” during a pre-game demo. Duped perhaps by the wobbly joystick and my unfamiliarity with the game, I had been fiddling for nothing, the victim of an illusion of control. And, indeed, when I started playing the game, I immediately noticed the difference. But for a while there, I was oblivious to my own irrelevance. I thought I was doing something that I really didn’t do at all.

The illusion of control is acute in our interactions with machines, as when we don’t know whether our push of an elevator button or a Coke machine selection lever 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 (say, 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?”).

The illusion that one has done something that one has not really done can also be produced through brute social influence, as illustrated in a study by Kassin and Kiechel (1996). These researchers falsely accused a series of participants in a laboratory reaction time task of damaging a computer by pressing the wrong key. All the participants were truly innocent and initially denied the charge, showing that they didn’t really *experience* damaging the computer. However, they were led later to remember having done it. A confederate of the experimenters claimed afterwards that she saw the participant hit the key or did not see the participant hit the key. Those whose “crime” was ostensibly witnessed became more likely to sign a confession (“I hit the ALT key and caused the program to crash. Data were lost.”), internalize guilt for the event, and even confabulate details in memory consistent with that belief—but only when the reaction time task was so fast that it made their error seem likely. We are not infallible sources of knowledge about our own actions and can

be duped into thinking we did things when events conspire to make us feel responsible.

Most of the things we do in everyday life seem to fall along the “normal” diagonal in figure 1.3. 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 automatism 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 to make one wonder 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 realize that conscious will is not inherent in action—there are actions that have it and actions that don’t.

The definition of will as an experience means that we are very likely to appreciate conscious will in ourselves because we are, of course, privy to our own experiences and are happy to yap about them all day. We have a bit more trouble appreciating conscious will in others sometimes, and have particular difficulty imagining the experience or exercise of conscious will in creatures to whom we do not attribute a conscious mind. Some people might say there is nothing quite like a human being if you want a conscious mind, of course, but others contend that certain non-human beings would qualify as having conscious minds. They might see conscious minds in dogs, cats, dolphins, other animals (the cute ones), certain robots or computers, very young children, perhaps spirits or other nonexistent beings.<sup>5</sup> In any event, the conscious mind is the place where will happens, and there is no way to learn whether an action has been consciously willed without somehow trying to access that mind’s experience of the action. We have the best evidence of an experience of conscious will in ourselves, and the second-best evidence becomes available when others communicate their experience of will to us in language (“I did it!”).

5. Dennett (1996) discusses the problem of other minds very elegantly in *Kinds of Minds*.

*The Force of Conscious Will*

Will is not only an experience; it is 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 instance, 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.<sup>6</sup>

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 backyard, 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 (Charleton 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 described it in 1812, "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" (298). At the extreme, of course, this view of the will makes

6. Wittgenstein (1974) complained about defining will in terms of experience, feeling that this by itself was incomplete. He noted that there is more to will than merely the feeling of doing: "The will can't be a phenomenon, for whatever phenomenon you take is something that *simply happens*, something we undergo, not something we *do*. . . . Look at your arm and move it and you will experience this very vividly: 'You aren't observing it moving itself, you aren't having an experience—not just an experience, anyway—you're *doing* something'" (144).

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 can't tell what God is going to do, we can't predict what the will is likely to do.<sup>7</sup>

The notion that will is a force residing in a person results in a further problem. Hume (1739) remarked on this problem 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 can't *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.

In the same sense, causation can't be a property of a person's conscious intention. You can't *see* your conscious intention causing an action but can only infer this from the constant relation between intention and action. Normally, when you intend things, they happen. Hume remarked in *A Treatise on Human Nature* (1739) that the "constant union" and "inference of the mind" that establish causality in physical events must also give rise to causality in "actions of the mind." He said, "Some have asserted . . . that we feel an energy, or power, in our own mind. . . . But to convince us how fallacious this reasoning is, we need only consider . . . that the will being here consider'd as a cause, has no more a discoverable

7. The idea that the will is a property that can vary in quantity, and that it inheres in people and has some natural force that can produce actions, nonetheless remains a useful part of some theories in modern scientific psychology (Muraven, Tice, and Baumeister 1998).

connexion with its effects, than any material cause has with its proper effect. . . . In short, the actions of the mind are, in this respect, the same with those of matter. We perceive only their constant conjunction; nor can we ever reason beyond it. No internal impression has an apparent energy, more than external objects have” (400–401). 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 concept 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, because 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 actual degree of constant conjunction 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). But the precise causal understanding of the conscious will that is captured in such discussions is not something that is linked in any direct way to the person’s experience of will.

The experience of will is merely a feeling that occurs to a person. It is to action as the experience of pain is to the bodily changes that result from painful stimulation, or as the experience of emotion is to the bodily changes associated with emotion. The person’s feeling of pain is not the same as the degree of twist applied to a person’s arm, nor is the person’s feeling of fear the same as the pattern of excitation occurring in the brain. In each case, the *experience* is incommensurable with the *occurrence*. An illusory pain is still pain, in an important sense, but it may not indicate damage at the location it signals; it may not be the *effect* of an injury at the site of *apparent* injury. Similarly, a conscious willing is still a

conscious willing even when it is illusory in much the same sense: it may not be the *cause* of an action of which it is the *apparent* cause. A person's feeling of will, and the associated report of this experience to others, is a key criterion commonly used for assessing whether conscious will has operated, but we must remember that this feeling is not the same as an empirically verifiable occurrence of mental causation.

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 actions, 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. If the empirical will were the measured causal influence of an automobile's engine on its speed, in other words, the phenomenal will might best be understood as the speedometer reading. And as many of us have tried to explain to at least one police officer, speedometer readings can be wrong.

### Mind Perception

Why would people mistake the experience of will for an actual 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 one particular 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 can't 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. Conscious will fits a larger conception in exactly this way—our understanding of *causal agents*. The intuitive experience of consciously willing our actions is something we return to again and again as we continue to assume that the experience of will reveals the force that creates

our acts, mainly because we have a more general understanding of causal agency that allows this to make sense.

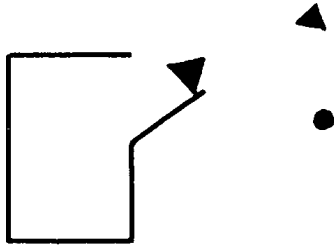
### *Causal Agency*

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 system. Dogs often seem to be goal-oriented; they behave in ways that seem to be understandable only 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 something we attribute just to living things; 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) 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.

This idea was illustrated in a classic study by Heider and Simmel (1944) in which people were asked to comment on a cartoon film of the motions of geometric objects—a big triangle *T*, a little triangle *t*, and a little circle *c*—around a box with an opening in it (fig. 1.4). The objects moved in the film in such a way that people almost always described *T* as chasing *t* and *c* around a house. People did not report any of the physical





**Figure 1.4**

A frame from Heider and Simmel's film of the movements of a dysfunctional family of geometric figures. From Heider and Simmel (1944). Courtesy Board of Trustees of the University of Illinois.

forces or interactions one might expect if these items were apprehended as physical objects. The objects weren't dropping or bumping or colliding but rather chasing and following and seeking. Apparently, the perception of causal agency can displace the usual way we have of perceiving physical objects, given the right circumstances.<sup>8</sup>

Of course, people are not simple geometric objects bouncing around a screen, and they have a considerable advantage over dogs, too. Their consciousness allows them the luxury of some insight into their own causal agency. People have access to an intricate mental dashboard display of cues regarding their goals because lots of cues to their agency appear in their thoughts and words. For this reason the inner workings of their causal agency can be interpreted in great depth. We may often learn from people what they think in advance of their actions, and we occasionally have this information available for ourselves as well, so we can construct elaborate understandings of likely actions and goals. The conscious causal agency of human beings is accompanied, in particular,

8. Heider (1958) went on to analyze the features of the perception of personal causation, which eventually gave rise to the field of attribution theory in social psychology (Gilbert 1998; Jones et al. 1972; Taylor and Fiske 1978). Unfortunately, this analysis failed to capitalize on the idea that causal agents are perceived as goal-oriented, and instead persevered on the notion that causal agents are perceived as performing behaviors that are themselves caused in some way (by factors that are either internal or external to the person). This led to immense confusion in the social psychological study of how people explain active, goal-seeking behavior (see the helpful correctives by Gilbert 1998; Kruglanski 1975; Malle 1999).

by relevant *intentions*, *beliefs*, *desires*, and *plans*. Let's consider each of these for a moment, even though they are rather elementary, just to remind ourselves how causal agency looks from the inside of the conscious causal agent.

*Intention* is normally understood as an idea of what one is going to do that appears in consciousness just before one does it. This is the thought that people usually associate most strongly with causing the action. Consider, for example, the case of watering the plants. If we actually remember to do this, the act typically seems like something we had the intention to do. In a study by Malle and Knobe (1997), people who rated the intentionality of twenty different behaviors put watering the plants at the top of the list along with such things as inviting a friend to lunch and refusing a salesman's offer. Behaviors such as sweating and yawning were rated as far less intentional. Although perceiving the causal agency of a dog might involve only figuring out what its goal might be, perceiving the causal agency of a person offers the additional prospect of considering what the person reports the goal might be—his or her professed intention. And perceiving causal agency in oneself involves coming to an understanding of one's actions in light of one's own conscious intentions.

We usually expect that a person who waters the plants has yet other conscious thoughts that are relevant to the watering. This person seems to have *beliefs* ("The plants need water," "This is water in my watering can"), for example, because it usually takes some understanding of the world and of the nature of the action to completely characterize what is going to be done. Such world knowledge seems particularly necessary for doing things that are extended or consequential (Danto 1963; Goldman 1970; Melden 1961). It doesn't take much in the way of knowledge about the world to wiggle your ears or bend a finger, but you must have a set of relevant beliefs and understandings about how your basic bodily movements will yield more distant effects if you want to do anything beyond wiggling and bending. Basic bodily actions thus seem to be nested within higher-level actions ("I bent my arm" is part of "I tilted the watering can," which is part of "I watered the plant"), even though all of them may be happening at once. We can describe an action at any of these different levels (Vallacher and Wegner 1985; 1987). However, at

each higher level, more needs to be known and believed about the context in which the basic bodily movement is occurring for that action to be understood. The more we can imagine that an agent knows or believes about the world, the more extended and involved are the actions and goals we can imagine the agent pursuing.

As a rule, the perception of human causal agency also involves understanding the person's *desire* ("I'm watering these plants in hopes of winning the All Green Thumbs Award again this year"). Desires are not always the same as intentions because they are sometimes descriptive of future circumstances that cannot be fulfilled with this act alone. In desiring a gardening award, for instance, one may intend to water the plants, but getting the award will also depend on other things such as competitors' plants, judges' decisions, and so on. In other cases, however, desire and intention seem to be the same, in that a person might conceivably want to water the plants only for the sheer fact of watering those plants. Still, it doesn't make much sense to attribute agency for an action to a conscious mind that doesn't want *something*, and for this reason philosophers of action seem to agree that a mental representation of one's desire is a key feature of the conscious willing of action (Anscombe 1957; Davidson 1963).

One other category of thoughts relevant to conscious will also can be important: *plans*. In a way, a plan is simply an intention that appears in mind some significant interval before the action. Searle (1983), for example, distinguishes between "prior intention" and "intention in action," and Bratman (1984; 1987) makes a similar distinction. Such prior intentions, or plans, are not usually seen as having the same causal role in our actions that immediate intentions have. I have a good idea, for example, that I will be going to Hawaii with my family next month (Oh boy!), and barring a calamity this is probably what I will do. And it is reassuring to know in this regard that there is a large research literature showing that people often do what they plan (Ajzen 1991; Gollwitzer 1993; Miller, Galanter, and Pribram 1960; Schank and Abelson 1977). Plans go awry, however. My cautious hedging about the upcoming Hawaii trip indicates that even the best laid plans . . . well you know the rest. Conscious plans to act do not seem to *compel* behavior in the sense that once the plan is in place the behavior always occurs. Instead, plans

just prepare the way. Often, they involve detailed specifications of how to act—the means or subacts of an action—and they thus make it more likely that when a situation arises in which the behavior is appropriate, the behavior can occur successfully.

Intentions that occur just prior to action, in contrast, do seem to compel the action. This is the basic phenomenon of the experience of will. In everyday language, of course, an action is often described as intended either when it is consciously planned (and we need not be conscious of it when we do it) or when we are consciously aware of doing it at the time we do it (and it might not have been planned in advance). Both are understood as acting on the basis of conscious will. However, if there is a conflict between them—as when we plan to stick to a diet at dinner but then end up (alas, all too consciously) splurging on dessert—the conscious idea of action that occurs just as the action starts is the one we will identify as our “true” intention. Prior plans that we fail to follow are then relegated to the recycling bin of false or mistaken intentions. The thoughts about the action at the time of action are the ones that prompt the strongest belief that we are causal agents who have made the act occur.

All these mental contents that seem to accompany and support causal agency in human beings need not be conscious at the moment of action. Instead, it seems that only the intention needs to appear in consciousness just as we act, whereas the beliefs, desires, and plans that may serve as the scaffolding for the intention need not be in consciousness. These other thoughts do seem to add substance to the idea that there is a conscious will that causes the action, and each of these kinds of thoughts seems to play a role in causing the action. But the conscious intention is, in a way, the mind’s “call” for the action it will do, and so the intention seems to be most immediately involved in the causation of the action.

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 others, 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. People 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.

### *Mechanisms and Minds*

We all know a lot about agents and goals, 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 instance, 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 behave*—doesn't fit very well with the physical causal relations we perceive all around us. Imagine 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 don't 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 doesn't 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.<sup>9</sup> Mechanistic explanation, in turn, is just splendid for understanding rocks and beer cans, and the movements of the planets, but it leaves much wanting 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 causes and effects as events in the world. In the mental explanatory system, 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

9. This odd possibility is the extreme consequence of attributing minds to things that can't talk. Chalmers (1996) makes the case for this theory, such as it is.

thinking, all the psychological trappings are unnecessary: A physical system such as a clock, for instance, doesn't 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, and 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 instance, 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 the juxtaposition of two findings by Jean Piaget: Children often neglect intention in making moral judgments, and yet they sometimes overattribute intention to inanimate objects. In the case of moral judgment, Piaget (1932) found that children before the age of seven or eight who are asked to decide whether a person has done something wrong don't concern themselves very much with the person's intention and focus instead on the damage caused by the action. For instance, in deciding how bad Haley was when she pushed Kelsey into the creek, a young child (say, aged six) might focus not on whether the pushing was done on purpose but rather on whether Kelsey's shoes were ruined by mud.<sup>10</sup> This is a bit odd because focusing on intentions could be very useful to children, particularly when claiming good intentions might reduce their punishment ("I pushed her in the creek to prevent her from getting heatstroke"). And while children do pay a bit more regard to their own intentions than to those of others (Keasey 1977), they still focus mostly on the damage. This lack of interest in intentions in moral judgments leads one to suspect that young children also may not appreciate minds in the same way grownups do.

In looking at how children judge inanimate objects, Piaget (1929) noted that they sometimes ascribe the properties of living beings, including the property of intention, to nonliving things. Based on the discussion

10. Research following up Piaget's initial suggestions has pointed out some problems with his approach and has suggested that the use of intention information in moral judgment comes somewhat earlier than Piaget had suggested, but it has also generally substantiated the conclusion that young children underplay the importance of intention in moral judgment (Schultz 1980).

of animism in anthropology (the tendency to ascribe living properties to nonliving things; Lévy-Bruhl 1910; Mead 1932), Piaget discovered that children could fall prey to the same thing—overattributing mental properties to systems that are better understood as mechanical. His attribution of animism to children has turned out to be controversial because he overstated the case for older children in contemporary cultures (Looft and Bartz 1969). Still, there is compelling documentation of animistic thinking in young children around the world. An interview with one four-year-old boy about why a toy boat floats, for instance, went like this:

- *Why does it not go to the bottom?*
- Can I make it go down to the bottom?
- *Try it and you will see.*
- It comes up again!
- *Why does it not stay at the bottom?*
- Because the man who is under this [under the roof] doesn't want to go down.
- *Here's a nail.*
- It will go to the bottom.
- *Why?*
- Because there is a man in here [in the nail] and he likes to go to the bottom. (Laurendeau and Pinard 1962, 209)

At first blush, Piaget's pair of insights suggest paradoxically that children underuse intention (in moral judgment) and also overuse it (in perceiving inanimate causality). This makes sense, though, when we realize that the child's notion of a mind is under construction. Without a fully developed idea of mental processes, children can fail to attribute intent when they should (in judging human beings) and attribute it too often when they shouldn't (in judging objects). Children are faced with the problem of building a picture of their own minds and the minds of others, and of achieving an understanding of what it is *not* to have a mind as well. Early in life, they guess that things without minds might have mind-like properties of intention and that things they will later learn *have* minds might not possess such intention.

Piaget's perspective has culminated in the contemporary literature on the development of theory of mind in animals (Premack and Woodruff 1978) and in children (Astington 1993; Perner 1991b; Wellman 1990),

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 (Astington, Harris, and Olson 1988; Carey 1996; Gelman 1990; Gelman, Durgin, and Kaufman 1995; Wellman and Gelman 1992). Neither the perception of the physical world nor the perception of the mental world is a “given” to the human newborn. Although the neonate has rudimentary abilities in both areas, both systems must be developed over time and experience as ways of understanding what is going on.

The field of psychology itself has noticed that different systems of thinking seemed to be necessary for understanding mind and matter. The main preoccupation of much of psychology in the twentieth century was translating mind talk into mechanism talk on the assumption that the two were entirely interchangeable. A telling quote from Donald Hebb (1946) on how psychologists should understand chimpanzees highlights what happened as a result:

A thoroughgoing attempt to avoid anthropomorphic description in the study of temperament was made over a two-year period at the Yerkes laboratories. . . . All that resulted was an almost endless series of specific acts in which no order or meaning could be found. On the other hand, by the use of frankly anthropomorphic concepts of emotion and attitude one could quickly and easily describe the peculiarities of individual animals, and with this information a newcomer to the staff could handle the animals as he could not safely otherwise. Whatever the anthropomorphic terminology may seem to imply about conscious states in the chimpanzee, it provides an *intelligible and practical guide to behavior*. (88)

This realization suggested to Hebb and others that the earnest project of eliminating mind entirely from the scientific explanation of behavior (Bentley 1944; Werner 1940) was misguided. You have to think about the animals’ minds in order to keep from getting mugged by them. A mental system for understanding even chimp behavior seems highly preferable to a mechanical system.

Perceiving mind and causal agency is a significant human ability. It is possible that this achievement is accomplished by a fairly narrow mental module, a special-skill unit of mind that does only this, and that in different individuals this module can thus be particularly healthy, damaged, or even nonfunctional. Leslie (1994) has called this set of skills a Theory-of-Mind-Mechanism (ToMM), and Baron-Cohen (1995) has proposed that such a mechanism may be injured or missing in some forms of



autism. He suggests that each of us has an “intentionality detector” that does the job of looking for actions that seem to be willed, in both self and others. The absence of this detector leaves us looking for physical or mechanistic explanations when psychological ones would really be better. Baron-Cohen has documented the “mindblindness” of autistic individuals in some detail, suggesting just how difficult life can be if one doesn’t have a quick and natural ability to comprehend other people’s minds. An example comes from Kanner’s (1943, 232) description of an autistic child: “On a crowded beach he would walk straight toward his goal irrespective of whether this involved walking over newspapers, hands, feet, or torsos, much to the discomfiture of their owners. His mother was careful to point out that he did not intentionally deviate from his course in order to walk on others, but neither did he make the slightest attempt to avoid them. It was as if he did not distinguish people from things, or at least did not concern himself about the distinction.”<sup>11</sup>

The idea that mind perception is variable has also been noted by philosophers. Daniel Dennett (1987; 1996) has 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, so that under some circumstances we might see our pet pooch as fully conscious and masterfully deciding just where it would be good to scratch himself, whereas under other circumstances we might 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, then, can depend on a host of factors and can influence dramatically how we go about making sense of the event. And making sense of our own

11. Oliver Sacks (1994) has documented the intriguing details of a life without mind perception. He recounts his interviews with Temple Grandin, an astonishing adult with autism who also holds a Ph.D. in agricultural science and works as a teacher and researcher at Colorado State University. Her attempts to understand human events—even though she lacks the natural ability to pick up the nuances of human actions, plans, and emotions—impressively illustrate the unusual skill most people have in this area yet take for granted. Grandin has cultivated this ability only through special effort and some clever tricks of observation.

minds as mentally causal systems—conscious agents—includes accepting our feelings of conscious will as authentic.

We're now getting close to a basic principle about the illusion of conscious will. Think of it in terms of lenses. If each person has two general lenses through which to view causality—a mechanical causality lens for objects and a mental causality lens for agents—it is possible that the mental one *blurs* what the person might otherwise see with the mechanical one. The illusion of conscious will may be a misapprehension of the mechanistic causal relations underlying our own behavior that comes from looking at ourselves by means of a mental explanatory system. We don't see our own gears turning because we're busy reading our minds.

### The Illusion Exposed

Philosophers and psychologists have spent lifetimes thinking about how to reconcile conscious will with mechanistic causation. This problem—broached in various ways as the mind/body problem, free will vs. determinism, mental vs. physical causation, and the analysis of reasons vs. causes—has generated a literature that is immense, rich, and shocking in its inconclusiveness (Dennett 1984; Double 1991; Earman 1986; Hook 1965; MacKay 1967; Uleman 1989). What to do? The solution explored in this book involves recognizing that the distinction between mental and mechanical explanations is something that concerns everyone, not only philosophers and psychologists. The tendency to view the world in *both* ways, each as necessary, is what has created in us two largely incompatible ways of thinking. When we apply mental explanations to our own behavior-causation mechanisms, we fall prey to the impression that our conscious will causes our actions. The fact is, we find it enormously seductive to think of ourselves as having minds, and so we are drawn into an intuitive appreciation of our own conscious will.

Think for a minute about the nature of illusions. Any magician will tell you that the key to creating a successful illusion is to make “magic” the easiest, most immediate way to explain what is really a mundane event. Harold 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 and birds, handkerchiefs, rabbits, and canes appear from nothing or disappear or turn into each other and back again.

The real sequence is often more complicated than the perceived sequence, 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 thought, preparation, and effort on the magician's part. The lovely assistant in a gossamer gown apparently floating effortlessly on her back during the levitation illusion is in fact being held up by a 600-pound pneumatic lift hidden behind a specially rigged curtain. It is the very simplicity of the illusory sequence, the shorthand summary that hides the magician's toil, that makes the trick so compelling. The lady levitates. The illusion of conscious will occurs in much the same way.

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. 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 don't see this. Instead, we readily accept a far easier explanation of our behavior: We intended to do it, so we did it.

The science fiction writer Arthur C. Clarke (1973) remarked that "any sufficiently advanced technology is indistinguishable from magic" (21). Clarke was referring to the fantastic inventions we might discover in the future or in our travels to advanced civilizations. However, the insight also applies to self-perception. When we turn our attention to our own minds, we are faced with trying to understand an unimaginably advanced technology. 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. 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 is a system that produces *appearances* for its owner. Things appear silver, for example, or they appear to have little windows, or they appear to fly, as the result of how the mind produces experience. And if the mind can make us “experience” an airplane, why couldn’t it produce an experience of *itself* that leads us to think that it causes its own actions? The mind creates this continuous illusion; it really doesn’t *know* what causes its own actions. Whatever empirical will there is rumbling along in the engine room—an actual relation between thought and action—might in fact be totally inscrutable to the driver of the machine (the 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 can’t ever know itself well enough to be able to say what the causes of its actions are. To quote Spinoza in *The Ethics* (1677), “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” (pt. II, 105). In the more contemporary phrasing of Marvin Minsky (1985); “None of us enjoys the thought that what we do depends on processes we do not know; we prefer to attribute our choices to *volition, will, or self-control*. . . . Perhaps it would be more honest to say, ‘*My decision was determined by internal forces I do not understand*’” (306).