Are We Free?

*Psychology and Free Will*

Edited by
John Baer
James C. Kaufman
Roy F. Baumeister

OXFORD UNIVERSITY PRESS
2008
Free will is not a topic that much occupies psychologists' thoughts or writings. Yes, we may reflect on the topic now and then in a desultory manner, but it hardly weighs heavily on the minds of modern researchers. The American Psychological Association, in collaboration with Oxford University Press, produced an eight-volume *Encyclopedia of Psychology* with many hundreds of entries. The volumes covered seemingly every topic under the psychological sun, running the gamut from *archetypes* and *altruism* near the beginning, to *voyeurism* and *xenophobia* toward the end. However, there was no entry on free will. In fact, *free will* did not even appear in the index as a term mentioned once, anywhere, in all eight volumes! (*Free association* was mentioned a number of times, though.)

Checking various other authoritative sources, we found much the same result. Occasionally an author would agree with the notion that research psychologists assume there is no such thing as free will. After all, why study behavior and look for regularities and laws if people can just do what they please any time they want, and in any situation? That's a good question, so most experimental psychologists probably think about free will for a few milliseconds, and then they go back to designing their next experiment. Psychologists prefer to leave the topic of free will to philosophers.

The editors of this book challenge psychologists to think hard about the topic, even if they have not done so since that introductory philosophy course many years ago. The challenge is perfectly appropriate, even if the answers in this volume may go largely unread by practicing psychologists (sorry to relate this news, if someone thought otherwise). Still, we agree with the premise
that those of us who study human behavior should think hard about free will. Perhaps the time when most of us think about this topic is when an experiment completely bombs. When subjects behave nothing at all as our precious theories and hypotheses say they should behave in a particular situation, the confounded psychologist may then come to believe in free will.

Does it really matter if free will exists? Human behavior is incredibly, overwhelmingly, complex. Edward O. Wilson in *Consilience* (1998) commented that the social sciences "are inherently far more difficult than physics and chemistry, and as a result they, not physics and chemistry, should be called the hard sciences." The determinants are so many that even if humans exert free will now and again, how would we ever know? Let us review some determinants here, but we will hardly be exhaustive. Starting with our genetic roots, even if we go back merely to the year 1500 or thereabouts, each of us has over 2 million ancestors. The gigantic genetic stew that each of us represents since then—happenstance meetings and couplings over the last 500 years—has led us to become the unique human beings we are today. The human genome is just beginning to be understood, but modern molecular behavioral genetics (agreeing with decades of twin studies) indicate that a good portion of human characteristics and behavior is genetically determined (in some cases, such as eyesight) and in other cases at least genetically influenced (scores on intelligence tests). After conception, all sorts of prenatal factors (what the mother eats and drinks, her hormone levels, her health or illnesses, her ingestion of drugs, and so on) affect the child who is born. And that child is born into cultures and circumstances that vary incredibly across the world. The culture and society in which we develop determine the language we speak, the food we eat, the behaviors we learn as appropriate for particular circumstances, among many other things. Anthropologists have long noted the huge influence of culture on behavior, although psychologists have come to study this topic only in the past 10 or 15 years.

Of course, just sticking with the factors noted in the previous paragraph, most psychological studies do not routinely consider genetics (not yet; it's coming, though), nor do prenatal factors occupy many researchers. Most psychological studies occur in Western cultures and most researchers don't give a thought about, say, communicating instructions for an experiment through the dominant language and simply assuming that they will be understood. Yet even within a relatively homogeneous culture, people's experiences differ greatly. Our parents, grandparents, teachers and schools, religious traditions and instruction, communities, and peers all help shape us as we grow. So do the books we read and the movies and TV shows we see.

The myriad factors listed in the previous paragraphs only begin to scratch the surface of determinants of human behavior. No wonder human behavior, even of the simplest sorts, is so hard to predict. No wonder so many differences exist among people treated the same way in the same conditions of our experiments. Our "error variance" is due not just to measurement error, but also to the
hundreds and thousands of differences among people in so many characteristics. Given these complexities, how could we know if free will does exist? Might it be only one more factor of difference in the bubbling cauldron of factors that help shape behavior?

FREE WILL: SOME PRELIMINARY THOUGHTS

We might profitably stop the chapter with the last sentence and let the reader move on. However, we do think that empirical psychology has several lines of research that might help shed light on issues surrounding free will. The astute reader will note that we have up until now dodged defining the main term of interest. We will borrow for our purposes the definition provided by Wikipedia ("Free Will," 2006): "The problem of free will is the problem of whether human beings exercise control over their own actions and decisions." This definition works for us, because psychologists have studied the topic of how humans control their own behavior. If we convert the problem of free will to the issue of control of behavior, psychologists may contribute to the conversation even if they still dodge the central issue in their studies—does free will exist?

Several research traditions are of interest in the study of the control of behavior, although we focus primarily on four: the response-choice paradigm developed by Benjamin Libet, the stop-signal paradigm developed by Gordon Logan and his colleagues, the process-dissociation procedure developed by Larry Jacoby and colleagues, and the free and forced report procedure developed by Asher Koriat and Morris Goldsmith. These paradigms are concerned with the issue of conscious control of behavior. Although the study of the voluntary control of behavior should not be equated with the problem of free will, at least its study may shed light on important questions. We review the paradigms and some main findings in the next two sections of the chapter and then discuss implications in a third section. The chapter ends with some general thoughts and conclusions.

NEURAL PRECURSORS OF ACTION

The control of behavior is a diverse topic that can be examined in myriad different ways. It is perhaps best to start our investigation with the simplest type of behavior control, the decision to make simple actions or to inhibit those same simple actions. Every moment of every day we act in many small ways, and even the most complicated action, such as giving a speech, consists of many smaller actions, such as deciding to emphasize a particular word or to inhibit another word. Some of the most interesting studies of the control of simple behavior are the neuropsychological studies of Benjamin Libet (1981). It was well known at
the time that Libet began his research that prior to motor movement there was an electrical change on the area of the scalp above the premotor cortex. This is known as the \textit{readiness potential}.

The readiness potential precedes movement by up to one second. Although this phenomenon had been thoroughly researched, Libet was the first to investigate the relation of the timing of the readiness potential to the conscious thought to make a motor movement. Libet set out to investigate when his subjects became consciously aware of an intention to act and whether the intention to act came before or after the readiness potential. In his experiments, subjects sat still while an electrode was attached to their scalp. They were then instructed to move their hand at random intervals throughout the experiment. He instructed his subjects to note the time when they first became conscious of the intention to move their hand. To measure the time accurately, Libet used a clock that was designed specifically for the experiment, with a hand that moved across the clock face at a speed that allowed subjects to precisely judge the time of occurrence of their intentions. Pilot tests to determine the accuracy of this measurement system were conducted by giving the subjects mild shocks and instructions to estimate the time at which they received those shocks. Surprisingly, subjects were accurate to within 50 ms.

Confident in the accuracy of his time measurement system, Libet began the experiments. Each time a subject made a hand movement, the readiness potential for the seconds prior to that movement was recorded and the subject was asked at what point on the "clock" he or she had first become aware of the intention to move his or her hand. The relative time of occurrence of the readiness potential, the conscious intention to move, and the hand movement could then be calculated. Most people believe that consciousness controls neural activation (thoughts) and movement (action), so they would believe that conscious thought to move the hand should come at the same time as, or slightly before, the readiness potential. This was not the outcome. The readiness potential preceded conscious awareness of intention by about 350 ms, on average (after correcting for error in determining awareness). The awareness of intention preceded action by 150 ms.

What does this tell us about free will? First, these data contradict the naïve view of free will—that conscious intention causes action. Clearly conscious intention cannot cause an action if a neural event that precedes and correlates with the action comes before conscious intention. This does not doom the concept of free will, however. Libet himself argued that a form of free will was supported by this data (Libet, 1999). Although conscious intention did not precede nor was it coincident with the onset of the readiness potential, it still preceded the hand motion by 150 ms. That leaves plenty of time for a person to inhibit the hand motion after they have become aware of the intention to move the hand. This indicated to Libet that while we may not have a free will, we do have a "free won’t"—we can stop behavior that has been initiated. Libet’s conception
of inhibition of a response as the locus of free will leads us to further examine research into inhibition.

Libet's research and ideas have been revolutionary in the fields of both philosophy and neuroscience. He has likewise attracted numerous critics. Below, we briefly detail recent relevant criticisms and related research. We note that far more criticism (Gomes, 2002; Klein, 2002; Oakley & Haggard, 2006; Pockett, 2004, 2006; Pollen, 2004) has been heaped upon Libet's work with the timing of perception (e.g., Libet et al., 1964) than upon his work on readiness potentials described above (for replies to the above criticisms, see Libet, 2002; Libet, 2006). Even one of Libet's most vocal critics considers his finding that readiness potentials precede conscious awareness of intention to be unassailable (Pockett, 2004).

We find the arguments of certain other researchers and philosophers, although less critical of Libet, to be more damaging to his finding. One group of researchers conducted experiments similar to those of Libet, but with subjects who were hypnotized to believe that they were not controlling their actions (Haggard, Cartledge, Dafydd, & Oakley, 2004). This group was compared to a group of subjects who knowingly controlled their actions and to a group of subjects who did not control their actions. The action being measured in all cases was moving a finger to depress a button. The subjects who controlled the button presses were instructed to press the button at random times, whereas the button depressed itself randomly for the subjects who did not control the button presses. Those who were hypnotized were told that the button would depress itself randomly, but they actually chose when to press it. All subjects rated each button press on a scale of how voluntary or involuntary the action was. They also estimated the time when their finger moved, allowing the experimenters to calculate error in perceived time of movement relative to the actual time of movement. Subjects who were hypnotized perceived their finger movements to be involuntary, just like those subjects whose finger movements were involuntary. Also, nonhypnotized subjects who made voluntary movements showed much greater anticipation of their movements, as shown by an earlier perceived time of action relative to the actual time of action, than did subjects who made involuntary movements. However, hypnotized subjects making voluntary movements showed the same amount of anticipation as those who made involuntary movements. This outcome indicates that the hypnotized subjects were not conscious of their free actions.

This research raises interesting questions that we cannot (for reasons of space) address here: If an action can be free when one believes it to be forced, is conscious volition irrelevant to the question of free will? This issue is related to Daniel Wegner's discovery that the feeling of self-control of behavior can be dissociated from actual causes of behavior. That is, people may cause an action but not be aware of it, or they may think they caused an action that was actually triggered by external forces (Wegner, 2003; also see his chapter in this book).
Whereas Wegner uses these findings to argue against the concept of free will, others have argued that even actions that stem from unconscious volition can represent evidence of free will, provided that the actions were not externally caused (Rosenthal, 2002).

Accepting that free will can be unconscious also defuses another criticism of Libet’s research: the idea that the “free won’t” would not be evidence of free will if preceded by an unconscious neural pattern similar to the readiness potential. If we accept that free will can be unconscious—although Libet (2006) himself refuses to do so—the possibility of an unconscious cause to inhibit action would not cause us to discard the concept of free will. These are thorny issues. We will leave to others the question of whether a conscious decision is necessary for free will or whether an unconscious decision can constitute free will. We do note, however, that the notion of unconscious volition is quite far removed from a straightforward conception of free will.

INHIBITION OF SIMPLE ACTIONS

The simplest method of investigating inhibition is the stop-signal paradigm, pioneered by Gordon Logan about 3 decades ago. In a standard stop-signal experiment, subjects perform repeated trials of a simple task (usually discriminating X from O). On a portion of the trials (usually around 20%), a tone is emitted at some point after the go stimulus (the X or O) has been presented but before the subject has responded to that stimulus (Logan, 1994). Subjects are instructed to stop performing the discrimination task (the go task) when they hear the tone (the stop signal). The go reaction time is the time from when the stimulus has been presented on a go trial until the subject makes a response. So in discriminating X from O, the time that the letter first appears on screen constitutes the beginning of the go task, and the time that elapses between the presentation of the letter and when the subject presses a key to respond constitutes the go-signal reaction time. The stop-signal reaction time is the time from when the stop signal was presented until the stopping process has finished. This time cannot be directly measured, because when the stopping process has ended nothing happens. The length of the stopping process can be estimated, however, by finding the area under the go reaction time distribution to the left of the probability of stopping (Logan, 1994). For example, if the probability of stopping on a stop-signal trial is 85%, and 85% of go-signal reaction times are less than 300 ms, then the time between the go signal and the stop signal is 300 ms. Subtracting the stop-signal delay gives the estimated stop-signal reaction time (Logan, 1994).

Researchers have investigated the effects of different tasks and stop-signal delays on stop-signal reaction time; they have also investigated subject population differences in the task. As we are using Libet’s work on the inhibition
of prepared responses (as determined by the presence of readiness potentials) as a justification for examining inhibition, we should point out that there has been some research looking at readiness potentials in the stop-signal paradigm (De Jong, Coles, Logan, & Gratton, 1990). This research found that, within the stop-signal paradigm, actions could be stopped even after a readiness potential had begun. This indicates that research on the stop-signal paradigm can inform the question of whether humans have “free won’t” and what the limits of that ability are.

Before describing other experiments, we should briefly explain a framework for looking at the data. In this framework, known as the horse-race model, there are two competing mental processes: the go process and the stop process. If the stop process is completed sooner than the go process, the action will be inhibited. Otherwise, the action will be performed. One key assumption of this model is that the two processes are independent. The data generally conform to this assumption (Logan, 1994). Following from this model, there are only four factors that determine whether a response will be stopped: the delay between the beginning of the go task stimulus and the stop signal (the stop-signal delay), the mean reaction time to complete the stop process, the mean reaction time to complete the go task, and the variance of the reaction time to complete the go task (the stop-process variance would also matter, but it is generally assumed to be zero for simplicity). To restate these factors, increased delay between the go signal and the stop signal, faster responses to the go task, and slower responses to the stop signal will lead to a decreased probability of stopping (Logan, 1994). Increased variability of the speed to complete the go task can either increase or decrease the probability of stopping. All experimental variables that influence the probability of stopping do so by affecting one or more of the above factors. We emphasize that the relative finishing times of the go and stop processes—not the relative starting times—determine whether or not an action will be stopped.

One of the first variables for which its relationship to stop-signal reaction time was tested was the delay between the presentation of the go stimulus and the stop signal. Data from an early experiment that tested this relationship are shown in the accompanying figure. These data show the measured reaction time to complete a go trial (left side of the figure) and the estimated stop-signal reaction time as a function of stop-signal delay (right side of the figure). Stop-signal reaction time declines as the stop-signal delay increases, and the go reaction time is longer than the stop-signal reaction time. That is, subjects are faster to respond to a stop signal than to complete the go task, and they respond more quickly to a stop signal as the delay between presentation of the go stimulus and the stop signal increases. This finding has been replicated (Logan, Cowan, & Davis, 1984).

Another consistent finding of stop-signal experiments is that subjects take about the same amount of time to inhibit a wide variety of different actions
TRIAL TYPE AND STOP SIGNAL DELAY

Figure 10.1.

(Logan & Cowan, 1984). People can inhibit both discrete tasks (such as discrete trials of discriminating among letters) and continuous tasks (such as typing or arm waving) with equal ease. Intersubject variability in stop-signal reaction times is also rather small (Logan & Cowan, 1984). Typically, most subjects can stop an action in about 200 ms.

Besides investigating the effects of various task parameters on stop-signal reaction times, researchers have also investigated the differences in stop-signal reaction times in different subject populations. There are numerous patient populations (e.g., those with frontal damage) that have a reduced ability to inhibit in daily living, so it is interesting to know if they show reduced ability to inhibit even in the simple stop-signal paradigm. This type of experiment has been performed with college students who score high in impulsivity (Logan, Schachar, & Tannock, 1997), children with attention deficit/hyperactivity disorder (ADHD) both on and off stimulant medication (Bedard et al., 2003), patients with frontal lobe damage (Dimitrov et al., 2003), and with healthy older adults (Kramer, Humphrey, Larish, Logan, & Strayer, 1994; Rush, Barch, & Braver, 2006).

College undergraduates who scored high on a measure of impulsivity showed a lengthened stop process relative to undergraduates who scored in the normal range on impulsivity (Logan et al., 1997). Because impulsivity did not
affect the response time to complete the go task, the effect is not due to a general slowing of cognitive processes. Children with ADHD had longer response times than normal children on stop trials, although they also had more errors on go trials. Those who were medicated with stimulant medication to treat the ADHD reacted more quickly and with fewer errors on both go trials and stop trials (Bedard et al., 2003). These data indicate that failure of inhibition is only one of the problems in children with ADHD. Patients with frontal lobe damage (specifically, frontal lobe lesions and frontal lobe dementia), although slower overall at performing stop and go tasks than control subjects, showed no differential impairment of inhibition in the stop-signal task (Dimitrov et al., 2003). This outcome seems surprising, given that frontal patients show reduced inhibition across many other tasks. Healthy older adults consistently exhibited slower stop-signal reaction times than younger adults (Rush et al., 2006). Although older adults were slower than younger adults on go trials, their slowing relative to younger adults on stop-signal trials was significantly greater than their slowing on go trials (Kramer et al., 1994), revealing a deficit in inhibitory control.

What can we conclude about the control of simple tasks through inhibition? Certainly, we can say that inhibition is a robust human ability. Humans can inhibit all manner of thoughts and actions (whether discrete or continuous). The time course of inhibition is also remarkably similar across different tasks and paradigms. Simply put, the ability to inhibit is one of the most important abilities that humans possess. Without strong inhibitory abilities, humans would be incapable of many of the most rudimentary forms of control over their own behavior (e.g., Hasher, Stoltzfus, Zacks, & Rypma, 1991).

ASSESSING CONTROL IN MEMORY PERFORMANCE

Although the above-mentioned findings highlight the role of inhibition in the control of human behavior, we must still keep in mind that these experimental paradigms take into account performance in only relatively simple tasks of very short duration. They do not directly speak to a much wider range of complex, deliberate actions that are typically longer in duration. Examples might include deciding what to say or not to say in a conversation, what answers to write in a classroom exam, or what information to recall as a witness in the courtroom.

Notice that in all of these types of situations, one can reasonably assume that individuals typically do not simply say or write everything that is on their minds. Rather, they make, or at least have the potential to make, responses in a conscious, deliberate manner, choosing what information to volunteer or withhold, depending on their circumstances and personal goals. The assumption that we have the capacity to carefully decide what and how much to say is especially critical when taking an oath to "tell the truth, the whole truth, and nothing
but the truth" in a court of law. Furthermore, by presupposing that individuals have control in deciding how to respond in these types of situations, one might further assume that whatever information comes to mind can vary in terms of the subjective experience associated with that information.

When trying to remember a past event, a person might be able to consciously recollect specific details regarding its context, or specific emotions that were experienced at the time, and be confident that this information stems from the occurrence of that specific event. Or perhaps certain details may unexpectedly come to mind that seem very familiar, but the person cannot consciously recollect the source of this information—and yet, he or she may still choose to recall or not to recall such details as part of the past event due to his or her sense of familiarity alone. In light of these possibilities, it is tempting to think that we have a great deal of control over our thoughts and actions in daily life, especially when it comes to remembering the past. But how much control do we really have under these types of circumstances to remember and communicate information accurately and completely?

At first glance, a cognitive psychologist would be the last person one would turn to for an answer to this question. After all, in laboratory settings, subjects are typically not allowed much leeway in their behavior. What types of responses a subject makes in an experiment are defined and carefully controlled by the experimenter. Responses that fall outside of the acceptable range are corrected in some fashion or considered outliers and subsequently ignored. However, in recent years we have seen a surge in interest in how to empirically study the influence of control processes in a wide range of cognitive tasks. Two influential paradigms concerned with the study of control in human memory performance that we will consider are the process-dissociation procedure, developed by Larry Jacoby and his colleagues, and the free and forced report procedure developed by Asher Koriat and Morris Goldsmith. Both experimental paradigms may be useful to the study of free will in that they provide quantitative estimates of the influence of cognitive control in a range of memory tasks. More importantly, research using these paradigms has helped to identify a number of variables that influence cognitive control.

The process-dissociation procedure (PDP) has been used to estimate the separate contributions of consciously controlled and nonconsciously controlled, or automatic, processes to performance on memory tasks (for a review, see Kelley & Jacoby, 2000). The process-dissociation procedure (Jacoby, 1991) is based on two theoretical assumptions. The first assumption is that performance on a cognitive task does not reflect the operation of a single mental process, but is rather the product of multiple processes operating conjointly. Second, the contributions of the processes to memory performance are independent, which is to say that they can be dissociated under certain experimental conditions or among subject populations. Indeed, the PDP approach directly challenges the notion that human beings have complete control over their behavior. For
the types of thoughts and actions that are typically analyzed using PDP, such as performance on memory-related tasks of recall, recognition, or completing word stems (e.g., ele______) under various instructional sets, one can, at best, only exert partial conscious control.

To actually implement PDP, one must use an opposition procedure (Jacoby, Woloshyn, & Kelley, 1989) in which the two types of mental processes (e.g., conscious vs. unconscious, controlled vs. automatic) that are presumably tapped by a given cognitive task are set in opposition to each other. For example, in one type of experiment, subjects might first study a list of words (e.g., element) and then perform a word-stem completion task (e.g., ele______). In one testing condition, the “inclusion” condition, subjects are asked to complete the word stems using only previously studied words (so ele______ would be completed by element if the subject is successful). In the second “exclusion” condition, subjects are told to complete stems with words that were not previously studied (so elephant, elegant, election, or electric, etc., would be appropriate). When subjects mistakenly complete a stem with a studied word in this exclusion condition—they produce element, contrary to instructions—their performance suggests the influence of automatic memory processes. That is, if production of element is elevated relative to a baseline condition in which the word had not been studied, the production must occur because the word was primed or activated by prior study but could not be successfully opposed or inhibited by conscious recollection of the event (a controlled process). The PDP, by assuming the independence of controlled and automatic memory processes, is able to provide quantitative estimates of the separate contributions of the two process types based upon performance in both inclusion and exclusion conditions. Details of the computations are outside the scope of this chapter, but can be found in Jacoby, Torn, and Yonelinas (1993), among other places.

The PDP has shown, for instance, that divided attention at study significantly reduces conscious recollection, but leaves unconscious or automatic influences on memory unaffected (Jacoby et al., 1993). Along similar lines, Hay and Jacoby (1996) used a variant of PDP to study the effects of experimentally trained habitual behavior and conscious recollection on cued recall performance. In an initial training phase, subjects studied words that were paired with typical responses 75% of the time (e.g., knee-bend) and atypical responses 25% of the time (e.g., knee-bone). Next, subjects studied a list of word pairs, some of which included words paired with atypical associates (knee-bone, in this context). During a final test, subjects were shown one member of each word pair and a fragment of its associate (e.g., knee-b_n) with instructions to recall the dominant response from the first list.

Under these conditions, the authors assumed that correct recall of typical pairs (knee-bend) could be based on either conscious recollection for the list or on subjects’ trained habitual response (an automatic influence). Conversely, when subjects mistakenly recalled the habitual response after having studied
an atypical pair, the authors assumed that habit was the basis for the response. By further assuming that these two sources for responding were independent, Hay and Jacoby (1996) calculated separate estimates of the contributions of conscious recollection and automatic habit on cued-recall performance. They found that whereas varying the amount of initial training of habitual responses did not affect the level of contribution of conscious memory in cued recall, it did affect the estimates of habit. In fact, estimates of habit (automatic responding) directly corresponded to the amount of initial habit training. Conversely, varying list presentation rate during study and cued recall response time during the test affected estimates of consciously controlled recollection, but did not affect estimates of habit. More recently, Hay and Jacoby (1999) administered this procedure to college students and older adults and demonstrated that cued-recall performance differed between age groups only because older adults' ability to consciously recollect the study list was impaired. However, the contribution of habitual responding was age invariant.

Taken together, these and a host of related studies have provided support for the notion that different memory processes can make independent contributions to performance on a single task. More importantly, these results appear to affirmatively answer the main question of free will proposed in this chapter, which is whether human beings exercise control over their actions and decisions. How much control human beings can exercise on any given task is another question altogether. To the extent that free will is related the control of human behavior, the PDP offers a unique approach to the study of free will. By assuming that performance on any given cognitive task cannot be the result of a single, controlled thought process, the PDP approach suggests that whatever behaviors one might consider pure acts of free will are really the product of controlled and automatic process or conscious and nonconscious thoughts and actions. Most important behaviors as analyzed by the PDP reveal the behaviors to be partly automatic and partly under conscious control. If one embraces a notion of free will that can accommodate such "process impurity" in determining human behavior, then the PDP offers a powerful analytic tool for measuring the relative contribution of controlled or conscious processing as well as identifying the factors that influence our control.

MANIPULATING OPTIONS TO REPORT MEMORIES

Another highly influential experimental approach to the study of control of human behavior permits subjects a larger degree of personal control to regulate both the quantity and quality of their responses in cognitive tasks. In typical memory experiments, subjects are told how to respond (recall studied material in order or recall it in any order) and they are often told whether or not to guess. A newer line of research, pioneered by Koriat and Goldsmith (1994, 1996),
focuses on the role that report option, the decision to volunteer or withhold information, plays in determining performance on memory tests. The basic idea is to test people under various conditions in which they are encouraged or forced to guess with other procedures in which they can withhold answers if they are unsure about the response. In an initial study, Koriat and Goldsmith (1994) compared the use of free and forced report procedures on both the quantity of memories produced and their accuracy, and in both recall and recognition tasks. Subjects attempted to answer general knowledge questions in either a recall format (e.g., What was the name of the composer who wrote the Moonlight Sonata?) or a recognition format (e.g., In response to the last question, choose among the following alternatives: Beethoven, Bach, Tchaikovsky, Schumann, Brahms). In addition, subjects were given either free report (i.e., respond to whichever items you believe you can answer) or forced report (i.e., respond to each and every question) instructions. Such a procedure seems particularly relevant to the study of free will, and comparison of performance under the two instructional sets should prove informative. Free report gives subjects the freedom to choose which questions to answer and which information to provide as answers, whereas forced report performance does not permit such freedom but requires subjects to respond (so both conscious and automatic influences may be operating).

The results showed that the use of forced report had no effect on increasing the quantity of correct responses reported relative to free report. This is somewhat surprising, because one might have expected subjects to come up with additional correct answers in the forced report condition, either by producing answers that were below a threshold for recall or just by random guessing alone. By contrast, the use of free report enhanced memory accuracy (the proportion of total responses made that were correct) in both recall and recognition (Koriat & Goldsmith, 1994, Experiment 1). This pattern of results also extended to performance on recall and recognition of studied word lists, a more conventional set of laboratory-based tasks (Koriat & Goldsmith, 1994, Experiment 2).

What these and earlier studies highlight concerning control of responding is that people appear to have little control over the amount of information they can retrieve at any given time. Changing recall criteria by encouraging subjects to make more responses does not lead to increases in amount of information a person can correctly remember, relative to standard free recall instructions with a warning not to guess (Bousfield & Rosner, 1970; Roediger & Payne, 1985; Roediger, Srinivas, & Waddil, 1989). In other words, allowing subjects more or less control in their memory reporting does not affect how much accurate information they can recall. By contrast, people can exercise some control over the accuracy of the information they do choose to recall when they are permitted to pass or to omit erroneous answers.

More important, a third experiment by Koriat and Goldsmith (1994) tested whether varying motivation for response accuracy would lead to further
improvements in accuracy. While attempting to answer general knowledge questions with free report instructions, subjects were given either moderate or strong incentives to provide only correct responses. Under moderate incentive conditions, the reward for correct responding was equal to the penalty for incorrect responding. Under high incentive conditions, subjects risked being penalized several times more heavily for committing errors than for being correct. The authors found that accuracy could be improved with increased incentives, but that the boost in memory accuracy came at a cost to memory quantity, as fewer correct responses were made. Koriat and Goldsmith (1994) argued that this quantity-accuracy trade-off occurred because accuracy can be improved by withholding answers, but because the process of covertly screening potential responses, or monitoring effectiveness, is not perfect, a cost is incurred in terms of decreased quantity of responses.

In subsequent work, Koriat and Goldsmith (1996) examined how the additional factors of subjective confidence and monitoring effectiveness influence control of memory performance, and they proposed a theoretical framework to account for their collective findings. Whereas subjective confidence involves monitoring the accuracy of potential responses as they come to mind, monitoring effectiveness refers to one’s ability to discriminate between correct and incorrect answers. Subjects first attempted to answer general knowledge questions under forced report conditions in both tests of recall and recognition, and then they assigned confidence ratings to their answers. In the next phase, they attempted to answer the questions under free report conditions with either moderate or high accuracy incentive.

Just as in their previous work, Koriat and Goldsmith (1996, Experiment 1) observed a quantity-accuracy trade-off in that subjects could achieve a higher level of accuracy in their free recall and recognition of answers to general knowledge questions when given a stronger motivation to be accurate, but at the cost of withholding a greater number of correct answers. In addition, subjects tended to volunteer responses in accordance with their confidence in the accuracy of those responses. Koriat and Goldsmith interpreted this finding to suggest that when remembering past events under conditions of free report, individuals apply a control threshold that allows for the output of responses that have the highest subjective probability of being correct. If a response does not surpass this threshold, then it will be withheld. And, as already mentioned, memory accuracy can be improved by increasing the motivation for accuracy, because such an increase encourages a person to set a higher control threshold.

Of course, the idea that judgments are based on a response criterion is not new. For instance, signal detection theory (SDT; Green & Swets, 1966; Wixted & Stretch, 2004), which is often utilized to analyze recognition performance data, assumes that an individual sets a response criterion, calling items whose familiarity or strength exceeds a response threshold “old” and items whose familiarity falls below the response criterion “new.” However, the utility of SDT is generally
restricted to forced-report testing conditions in which a subject must respond either "old" or "new" to each and every test item. By contrast, the forced and free report procedure allows for the measurement of changes in response criteria that occur under both forced and free report conditions. Tulving (1983) also proposed that people have "conversion thresholds" that operate to determine whether they will report a retrieved event as a memory, with some conditions of responding leading to more stringent thresholds than others.

With regard to monitoring effectiveness, one can imagine that having a poor ability to distinguish between right and wrong answers would undermine a person's ability to improve accuracy in memory reporting. Indeed, when subjects attempted to answer "deceptive" general knowledge questions such as, "What is the capital of Australia?" (hint: the correct answer is not Sydney), monitoring effectiveness proved to be quite poor, and as a result, increased control of memory reporting yielded little or no benefit (Koriat & Goldsmith, 1996, Experiment 2). In other words, subjects continued to respond on the basis of their confidence in the accuracy of retrieved information, except that due to the deceptive nature of the questions, subjective confidence was no longer a reliable guide for accurate responding. For questions that were not deceptive, monitoring effectiveness improved, and as a consequence, greater increases in accuracy were observed at lower costs in memory quantity. In this case, subjective confidence served as a more reliable, albeit imperfect, response index (Koriat & Goldsmith, 1996, Experiment 2). The authors argue that, in theory, when monitoring effectiveness is perfect, there should be no quantity-accuracy trade-off whatsoever. (By the way, Canberra is the capital of Australia).

By contrast, SDT treats subjective confidence as being synonymous with memory strength in that confidence judgments are used to construct ROC (receiver operating characteristic) curves by assuming that confidence directly corresponds to memory strength. Thus, one advantage of Koriat and Goldsmith's theoretical framework is that it can take into account situations in which subjective experience is not a reliable basis for remembering. Consider the use of eyewitness testimony to identify suspected criminals from lineups. Eyewitnesses who are very confident in their identification can provide very compelling evidence in a court of law. And yet, eyewitness research has shown that subjective confidence may not be a reliable guide for accurate identification (for a brief review, see Wells, Olson, & Charman, 2002). For instance, providing reinforcement to eyewitnesses during in a lineup procedure (e.g., "Good job. You are a good witness") may increase their confidence in their response without improving its accuracy (Wells & Bradfield, 1999). Also, repeated questioning of eyewitnesses about mistaken memories does not lead them to revise their recollections. Instead, it leads eyewitnesses to inflate the confidence of their false recollections (Shaw, 1996; Shaw & McLure, 1996).

More recently, Goldsmith and Koriat (1999) have examined an additional means of subject control in memory reporting—control over "grain size," which
is the level of generality or detail of a response. For instance, in response to the question, "At what time did the robbery take place?"—rather than simply volunteering or withholding a response, an individual might choose to frame the answer in a way that is more likely to be accurate (e.g., "in the late morning" as opposed to "at 11:30 A.M."). At the same time, relying too much on accuracy may render the reported information uninformative, such as when a person is asked when World War II occurred, and, as a response offers, "Sometime in the 20th century." Indeed, research shows that the choice of grain size is influenced by a person's attempt to compromise between competing tendencies to be both accurate as well as informative in a given situation (Goldsmith & Koriat, 1999; Yaniv & Foster, 1995).

According to Koriat and Goldsmith, the control that individuals exhibit in their memory reporting may be represented by a control threshold for responding, the setting for which will vary depending upon a person's goals and circumstances. To quantitatively assess the factors that influence this control process, Koriat and Goldsmith (1996) developed an analytic procedure, the quantity-accuracy profile, or QAP. QAP describes the joint levels of quantity and accuracy performance that might be attained at various control thresholds, depending upon an individual's overall level of retention and monitoring effectiveness. In addition to providing standard quantity and accuracy measures or memory performance, the QAP provides estimates of monitoring effectiveness and control, and encourages researchers to take into consideration subjects' goals and incentives in experimental situations. To date, though, the method does not accommodate control over grain size in memory reporting.

Koriat and Goldsmith's (1996) framework has been applied to subject populations, such as children, older adults, and schizophrenia patients, who exhibit deficits in memory performance to determine the extent to which such deficits may be attributed to impaired control processes (Dunion et al., 2001; Kelley & Sahakyan, 2003; Jacoby et al., 2005; Koriat et al., 2001; Meade & Roediger, 2006; Rhodes & Kelley, 2005). For instance, it has been shown that children as young as 8 to 9 years old can take advantage of free report to increase the accuracy of their memory reports. However, such children could not achieve the level of performance, both in terms of quantity and accuracy, exhibited by children several years older (Koriat et al., 2001). Similarly, older adults show smaller gains in accuracy performance relative to younger adults when they shifted from forced to free report testing conditions (Jacoby, Bishara, Hossels, & Toth, 2005; Kelley & Sahakyan, 2003; Meade & Roediger, 2006; Rhodes & Kelley, 2005).

For our purposes, it is worth noting that the form of control that individuals appear to exercise in the free and forced report procedure is the capacity to suppress or inhibit already retrieved candidate responses at a relatively late processing stage, just prior to verbal report. Candidate responses may be retrieved through conscious effort, nonconscious or automatic processes, or both. The theoretical
framework of Koriat and Goldsmith does not assume that information can only be initially retrieved in a nonconscious or automatic manner. Indeed, Jacoby and colleagues have recently shown that people can control what information comes to mind during retrieval, a process they term early selection (e.g., Jacoby, Shimizu, Daniels, & Rhodes, 2005). By contrast, the role of the control mechanism in the Koriat and Goldsmith framework is to serve as a gatekeeper, after responses have been generated or selected. Above-threshold retrieved information is passively allowed to reach the stage of verbal report; sub-threshold retrieved information is suppressed or inhibited from output. This framework challenges the notion that human beings have the ability to exercise full conscious control of their behavior on memory-related task performance. Rather, Koriat and Goldsmith (1996) propose a more modest domain for the control of human memory.

CONCLUSIONS

We have reviewed four bodies of evidence that deal with the issue of conscious control of behavior. We have danced around the issue of whether conscious control is to be equated with free will; in fact, we suspect that at the most basic level, the answer must be no. Even behavior that subjects believe to be completely under conscious control is influenced by external factors, as discussed above. Yet even if we cannot draw firm conclusions about free will from the body of research that we have summarized, we do think that the four areas of research we have reviewed are relevant to the issues at hand. However, the work in each area of research may raise as many questions as it settles at this point. For example, the ability to inhibit responses is powerful and one could reasonably make the argument that without it, free will would not be possible, because we would not be able stop what some force (external or internal) seemed to impel us to do. Unfortunately, this still leaves unanswered the question of whether we have free will, despite our demonstrated abilities to consciously inhibit behavior. If we do assume that free will can be directly observed in inhibition of behavior, then is free will measured by inhibition?

The latter question is especially poignant with respect to those people who show marked inhibition deficits, such as children, the elderly, and certain patient populations. If we agree with Libet (1999) and consider inhibition of an unconsciously determined action to be equivalent to free will, then these populations suffer from impaired free will. At first glance, it seems quite odd to view free will as a measurable mental ability and to characterize broad swaths of the population as having impaired free will. Upon further consideration, though, we see that this view is already implicitly, if not explicitly, accepted in many fields, as well as (to some degree) enshrined in common wisdom about behavior. This principle is most visible in how the legal system treats offenders depending upon whether or not an act was premeditated. In the United States and in
most other countries, legal systems consider premeditated murder to be a much worse offense than murder that is not premeditated, even if the circumstances of the murder are otherwise identical. This accepted principle shows that society (or the legal system) considers people to have less free will when they are in a fit of rage than when they plan carefully and laboriously to commit an act.

Returning to the work of Kortat, Goldsmith, and colleagues, we can also ask similarly difficult questions. For instance, should children, older adults, or for that matter, any one of us, serve as witnesses and take oaths to tell “the truth, the whole truth, and nothing but the truth” if we fully recognize the difficulties of monitoring memory accuracy and the severe consequences for reporting false memories? Again, if the locus of free will is in the control of behavior, then those who can exercise less control of certain behaviors have, by this logic, less free will. If people show improved control in a task with practice, are we justified in saying that they have enhanced free will? Although this idea sounds bizarre when stated plainly, voluntary control as directly related to free will is a common concept. Belief in the ability of “self-control,” whether in resisting eating too much or driving too fast or drinking alcohol, is widespread. One reason people consult clinical psychologists is to help them gain self-control (in weight management, in alcohol consumption, or in managing their level of anxiety or depression). We have focused on approaches from experimental psychology in this chapter, but certainly issues in the clinical settings arising from the self-control exerted (or not exerted) by patients is another relevant arena in considering free will. The issue of free will must always be part of psychology, even if it is never completely settled.

REFERENCES


