CHAPTER 3

Self-Report Is Unreliable Because Cognition Is Often *Automatic*

Like other sciences, psychological science faces daunting methodological challenges. How can you study the very thing you wish to study if it is invisible? Over the centuries, other sciences have developed tools that we know and trust—microscopes, telescopes, calculus (and more complicated forms of mathematical reasoning), centrifuges, and so on—that allow scientists to study even the things that cannot be seen by the naked eye and that once existed only in theory. Psychology has two disadvantages relating to its tools relative to other sciences: (1) it is a young discipline, having only used scientific methods for about 100 years, with tools that are less developed and trusted relative to older sciences; and (2) the object of study is ourselves.

There is an old adage that when humans know their behavior is being studied, it changes the way that humans behave, so that "natural" responding is altered (the **Hawthorne effect**). This is complicated even further when studying social cognition as opposed to behavior. Methods for studying social cognition are not only complicated by people becoming unnatural when aware they are being studied but by people being *unable* to accurately report how they are thinking even if they wanted to. How can you tell what someone is thinking, what they intend, and how they produce the responses of which they are consciously aware if we rely on an unreliable conscious response as the method by which we learn these things? Psychological science is often disparaged with the quip "it's not rocket science." But, at least rocket scientists have the tools they need to do the job. Allport (1954) stated that "Civilized men have gained notable mastery over energy, matter, and inanimate nature generally, and are rapidly learning to control physical suffering and premature death. But by contrast we appear to be living in the Stone Age so far as our handling of human relationships is concerned. Our deficit in social knowledge seems to void at every step our progress in physical knowledge" (p. xiii). It is both a difficulty and an allure of social cognition that the tools that can penetrate the true nature of thinking needed to be developed.

Hawthorne effect: How people naturally behave is obscured by the knowledge that one's behavior is being studied, leading people to alter their behavior so that it does not reflect what is natural.

In the earliest days of psychological science, Wilhelm Wundt's lab (with students such as James McKeen Cattell, Oswald Külpe, and Edward B. Titchener) used **introspection** as a method for examining psychological processes. Unlike how it is used in everyday language, introspection as a tool for psychological discovery did not refer to casual processes of reflection. Instead it referred to extensive training in attending to elemental sensations and images experienced in the mind (Wundt, 1862). Through practice, over thousands of introspections, one could develop skill at self-reporting on low-level cognition: "perception, apperception, discrimination, judgment, and choice" (Boring, 1953, p. 172). Yet, within Wundt's own lab, Oswald Külpe came to conclude that this technique was ineffective for examining thought processes. He argued that one did not have access through introspection to examine the processes that direct thinking. Thus, one category of reasons to distrust self-report is a lack of ability for one to access the information with which we investigators are concerned. [1]

Introspection: A methodological tool that relies on self-report about one's own psychological experience, but a specific type of self-report that requires intense training in how to observe one's own perception, discriminations among stimulus features, judgment, and choice.

Another, entirely different reason that trust in self-report as a tool can erode is a lack of desire on the part of the individual to report to investigators the true nature of their experience, even if able to access it. Perhaps self-report is reliable if the task is to decide which of two lights shine more brightly; is it equally reliable when the task is to distinguish which of two people is more skilled? As the targets of our inquiry acquire social value, that value can deter us from wanting to share the truth. People can be motivated to lie when reporting on their social cognition. In some cases, the perceiver may make a conscious choice to lie during self-report—in other cases they may not even know they "lie." We turn next to more fully examining reasons to distrust self-report: decisions to deceive others, deceiving the self, and lacking the ability to access what to report.

STRATEGIC DECISIONS TO DECEIVE: STEREOTYPING AND LIES OF SOCIAL DESIRABILITY

Self-report is often fallible as a measurement tool because of **social desirability bias** (e.g., Paulhus, 1991; Schlenker & Leary, 1982). People have a desire for others to view them in a way that will contribute positively to self-esteem. This creates concerns about being seen negatively by others, an apprehension, or social anxiety, about being evaluated negatively that threatens to lower self-esteem (e.g., Stephan & Stephan, 1985). Such anxiety about looking bad, or a desire to look good, motivates people to make a positive impression when they have doubts that natural responding would do so. A strategic intention to deceive others with one's self-presentation can occur when people feel they are being observed or are the focus of attention. It also occurs when concerns are raised about how one's natural response corresponds with what others expect, such as expressing beliefs that do not conform with the norms and standards for what ought to be done that are held by a valued social group (e.g., saying something inappropriate or harassing that would lead others to think those statements stem from a biased person). When one worries about interpersonal failure, especially those that people find threatening to self-esteem because they imply one is an immoral person, social desirability can lead to deception.

Social desirability bias: Concerns about being viewed negatively by others and apprehension (social anxiety) about being evaluated negatively that threatens to lower self-esteem. It is a desire for others to see the self in a way that contributes positively to self-esteem, including strategic concerns about presenting the self to avoid being seen negatively.

Not Stereotyping as a Lie: The Case of Normative Pressure to Be Unbiased

To avoid public disapproval, people can be unwilling to express publicly a stereotype that they privately believe (e.g., Lambert, Cronen, Chasteen, & Lickel, 1996). They allow the lie to conceal an unflattering personal fact that they are unwilling to admit to others. This type of lie not only happens among friends but people lie to strangers. For example, people lie to pollsters and on self-report measures by reporting they support a normatively popular person or thing they actually dislike. Stereotypes often operate in this way. A specific version of this type of lie relates to political candidates from disenfranchised or minority groups that have popularity in the culture. To oppose such a person, even if for reasons that have nothing to do with being prejudiced, could appear to be due to prejudice. Thus, social desirability suggests saying you support the minority candidate, even if you do not. This has been called **the Bradley effect**, named after an African American mayor from Los Angeles who lost the race for governor of California in 1982 after polls overestimated his support. The accepted belief about why this happened is that White voters attempted to appear unbiased in what they told pollsters and news organizations. Succumbing to social desirability concerns, they claimed to support this African American candidate. What they reported was a lie in that they did not truly intend to vote for the minority candidate. In the privacy of a voting booth many did not.

The Bradley effect: When people lie on self-report measures by reporting that they support/approve of a normatively popular person or thing they actually dislike. Stereotypes often operate in this way. To avoid public disapproval, people can be unwilling to express publicly a stereotype that they privately believe.

In cases of stereotyping such as this, there is the potential to cause harm by one privately acting in ways that are the opposite of what one says. If many people do so, it creates an appearance that a problem does not exist, but private actions promote stereotype-guided behavior and disparities. Moving beyond elections, consider important jobs in which the public places trust: doctors, teachers, nurses, police officers, and so on. If such people publicly assert they have no bias against women, against Black men, or against transgender individuals, then it creates an illusion in the society at large that no disparities exist. However, if they privately endorse such stereotypes, it can lead to treating those individuals differently and perpetuating or extending disparities, with the cultural impact never being detected due to their explicit denouncement of bias.

Consider the example of sentencing decisions made by jurors (Glaser, Martin, & Kahn, 2015). Jurors claim to believe in an unbiased judicial system and to agree with the principle of "statutory (and ethical) irrelevance of race in the determination of suspicion, guilt, or punishment" (p. 2). However, past research has shown that race plays a role in how jurors actually mete out punishment for identical crimes. White defendants are treated more leniently for the same crime, and this effect is even found for the death penalty, where Black defendants are more likely to be sentenced to death, especially if the victim is White (e.g., Eberhardt, Davies, Purdie-Vaughns, & Johnson, 2006). Glaser et al. performed an experiment to examine this matter. Do jurors find a person guilty and susceptible to the death penalty differentially as a function of race, despite stating they would never use race as a factor in determining the decision to convict? Research participants were asked to decide whether to acquit or convict a suspect in a murder trial and were provided with a case summary that suggested guilt. The summary left it ambiguous as to whether the victim was Black or White, but manipulated whether the defendant was Black or White. When the jurors thought the maximum sentence was imprisonment, they did not treat a Black defendant any differently than a White defendant. The private decision about sentencing a person to prison made in the experiment matched what they publicly endorsed. However, when the maximum sentence was *death*, a bias appeared. They were reluctant to provide a

death sentence to a White defendant relative to a Black one—that is, they were more likely to convict a Black defendant who faced the possibility of the death penalty. They publicly say that justice, to them, is blind. But when a life is on the line, what they privately do is not matched to what they publicly say.

A similar break between what people publicly say and privately do is seen with research with police officers. White police officers make the claim that their training and experience makes them relatively immune to anti-Black biases that are displayed by laypeople who are White. On one stereotyping task, where a research participant is asked to simulate firing a weapon if they encounter a person holding a gun (as opposed to holding a wallet or a phone), participants fire more quickly if the person holding the gun is Black, and they make the decision to not fire more quickly if the person holding a wallet is White (e.g., Correll et al., 2007). Police officers show this exact same bias. And the bias is greatest in officers who work in areas with the highest concentrations of Black and minority citizens. In fact, officers trained to work on street violence and gang activity show the bias as much as laypeople (Sim, Correll, & Sadler, 2013). Where the experience they publicly tout as immunizing them from bias is greatest, private bias is greatest.

One final example comes from the medical community, where strong normative pressures create a need to provide the best possible care and to do so without prejudice, yet racist and sexist responding still prevails in private. Even when doctors explicitly proclaim to adhere to their Hippocratic oath, they treat patients from different social groups differently. Green et al. (2007) showed that when medical doctors reported feelings for Black patients that were equally positive to those of White patients, many nonetheless held private negative feelings toward Black patients (private feelings of bias were assessed using a tool we review in Chapter 4: the IAT). Do these hidden negative prejudices contribute to a bias in their treatment recommendations? The doctors of medicine (MDs) in this experiment were provided with a description of a 50-year-old man with chest pain, accompanied by an electrocardiogram implying he had anterior myocardial infarction. A picture of the patient was included as well, with half the MDs seeing a picture of a White man, the other half received a picture of a Black man. All other information was identical from doctor to doctor. Their findings showed that Black patients were diagnosed more often with coronary artery disease (CAD). However, the preferred treatment for CAD—thrombolytic drugs—was not recommended more often to Black patients. The more that the test of their private attitudes revealed prejudice in the MDs, the less likely they were to recommend thrombolysis for the Black patients. Public attitudes did not impact treatment recommendations. A preferred medical treatment was being systematically denied, but the doctors did not report having any bias at all in how they would treat patients. But differences exist.

One might argue that this is not the doctor yielding to a norm and hiding their "true" negative feelings but a case of the doctors simply not recognizing their negativity and truly believing they are unbiased (as is explored next in the section on aversive racism). This is possible. However, it is also possible that doctors are merely saying what is normative in public and are aware of their private prejudice. They lie. Evidence in support of this would be provided by doctors being more likely to show bias when they have greater social desirability concerns. For example, Wolsiefer et al. (2023) found that physicians with higher levels of social desirability concerns had higher levels of private anti-Hispanic bias.

Stereotyping as a Lie: The Case of Normative Pressure to Be Biased

We just reviewed cases where the normative pressure is to like someone (such as pressure to not stereotype a member of a different group), or to label someone as qualified and good,

while privately one feels otherwise. The inverse of this happens as well. In such instances stereotyping and bias emerge because people feel pressured to publicly express bias when privately they do not feel any. Imagine a person who is low in prejudice yet is living or working or studying among people who they believe to be high in prejudice. An unbiased person could express bias in these situations because they fear not fitting in. The pressure to acquiesce to bias is great. Many of us have felt this pressure when hearing a sexist or racist joke that we seemingly supported, allowing our fake approval to give legitimacy to the bias. To this day I am haunted by a shameful smile from my childhood to a terribly offensive joke told by a friend's father while giving me a ride home from nearby Manhattan as a favor to my family. I wanted to condemn him or jump out of the car and walk home. But we were 20 miles from home, and I was 15, in a car pool full of adults. All I could muster was a fake smile—a lie—as he and the other adults in the car laughed at a racist joke.

There is an entire literature on racist humor that explores this way that stereotyping can be a deceit exacted to gain social approval from people who, unfortunately, may be doing reprehensible things. Ford and Ferguson (2004) proposed a **prejudiced norm theory of humor**. In this theory, jokes are said to be used to communicate group norms and can be used as a way to signal agreement with the group (and as Ungson, 2019, suggested, they can also signal dissent from the group if jokes are used to challenge a norm). For example, when male research participants heard sexist jokes they then believed that a norm of anti-sexism was less strongly in place, and that being sexist was more acceptable (e.g., Ford, Boxer, Armstrong, & Edel, 2008). Some of the men then acted in a more sexist manner as a result, and for those who did not, a pressure to not dissent was felt. A joke is more than a joke, but an expression of a norm, and many men may find themselves in locker rooms and boardrooms where sexist norms are communicated through humor. Nonsexist men may decide to lie and not voice their opposition to sexism because of such social pressure. In keeping quiet they do more than give the false impression that everyone shares the sexist norm. Jokes are only humorous when we all agree they are benign, or done without malicious intent (e.g., Warren & McGraw, 2016). Thus, laughing at sexist jokes also communicates that sexism is more than just normative, but it is benign or harmless.

Prejudiced norm theory of humor: Theory that jokes are used to communicate group norms and can be used to signal agreement with the group (even if one privately disagrees). For example, when male participants heard sexist jokes they believed that the norms of anti-sexism were less strongly in place, and that being sexist was more acceptable.

It is not just in humor where people "lie" by stereotyping. Social pressure within one's circle (family, friends, carpoolers, work colleagues, teachers, fraternity brothers, teammates, etc.) can exist to pressure one to denounce certain groups that are disliked within that circle. Socially shared sets of beliefs are transmitted, and a good group member is supposed to share those negative beliefs. People may privately not agree with a stereotype but may publicly agree with it so they are seen as agreeing with their group (Katz & Braly, 1933). An example from politics would be if a valued social circle had a shared dislike of a candidate because of their race. In the opposite of the Bradley effect, one would now fail to publicly endorse a candidate who one privately supports out of concerns about social disapproval. In this case, polls underestimate the vote, leading a figure that polling suggests is trailing (or with a narrow lead) to win (or win by a landslide). Some believe this is why polls underpredicted the victory of Barack Obama in 2008.

Let us look at one last example of how we lie with a public statement that is misaligned with private beliefs. People will publicly say they disagree with a minority, when private measures show far more agreement than they self-report. People are reluctant to express public agreement with a negative group. At times this is a true reaction (there is actual power in numbers and we feel those in the minority are less correct). Yet at times this is a lie,

and we say we disagree with the minority even though the minority has convinced us to agree with their view—there is **minority influence**. Moscovici, Lage, and Naffrechoux (1969) conducted studies in which four research participants and two confederates (people who actually work for the research team pretending to be fellow participants) were asked to report the color of a series of slides. All were blue slides that varied in their light intensity. The two confederates (a numerical minority) were asked by the experimenter to give an incorrect response and report seeing the color green. The participants still reported seeing the color blue on 92% of their responses, thus mostly disagreeing with the minority that consistently reported seeing the color green. This was a lie because more subtle measures showed that the participants actually agreed with the minority far more often than 8% of the time and were just reluctant to say so publicly. When later brought into a private room and exposed to blue–green disks, over 35% of them reported seeing green.

Minority influence: The power of minorities (often defined numerically) to persuade people. Often, influence is initially private, with people reluctant to publicly express agreement with minorities (due to normative pressures to agree with the consensus). With time, the stereotype of minorities as wrong can fade within individuals and public opinions may shift.

Moscovici and Personnaz (1980) made this same point in an experiment that utilized chromatic afterimage. If an individual fixates on a white screen after focusing on a color, they will see the complementary color on the screen. Thus, if a participant actually perceives the slides to be green, they will be more likely to report seeing red (the complement of green) than others who actually perceive the slides as blue, who should report seeing yellow (the complement of blue) during the test of their chromatic afterimage. They found that when a minority reported seeing green, the participant's public report was that they saw blue. However, they saw red during the chromatic afterimage test with a greater frequency than research participants who were confronted with a majority who reported seeing blue on the initial task. While minorities have a small influence publicly, the influence is much larger in private, indicating that the public response was a deception (see also Nemeth, Swedlund, & Kanki, 1973).

Lying to Say What Is Socially Desirable Can Be Detected from Nonverbal Behavior

As just reviewed, when we have social desirability concerns, we may say and do things that will align us with what is "appropriate" even if we privately disagree. However, as shown above, researchers can detect such misalignments with cleverly designed experiments. One set of clever procedures takes advantage of the fact that when our true beliefs and feelings are misaligned with overtly expressed beliefs and feelings, we betray our hidden views with our *nonverbal behavior*. As we review in more detail later in this chapter, nonverbal behavior is more difficult to control than the things we consciously say. We may say we are not mad but our anger screams out in other ways. We may lie and say we have not broken the law but in doing so we may be nonverbally admitting to the crime. Researchers can use measures of nonverbal behavior to reveal true attitudes and beliefs despite what is being publicly said.

Dovidio, Kawakami, and Gaertner (2002) used this fact to explore how racism might be expressed through one's nonverbal behavior, even when one overtly reports having nonracist views and feelings. They argued that while consciously intended behavior is typically predicted quite well by publicly expressed beliefs, it is not predicted by private beliefs and attitudes. These more implicitly held beliefs and attitudes instead predict more subtle types of behaviors, such as nonverbal displays of bias. The fact that overt and explicit bias

is predicted by what people explicitly say is nicely illustrated in an experiment by Fazio, Jackson, Dunton, and Williams (1995). They looked at biased responding to a case of police brutality that had great publicity in the early 1990s (the beating of a Black man named Rodney King). Responses to this incident, including reactions to the subsequent trial of the officers and support for the Black community, were predicted by measures of the participant's publicly expressed racial prejudice (see also Dovidio, Kawakami, Johnson, Johnson, & Howard, 1997). However, privately measured attitudes did not predict these conscious responses. What is predicted by privately measured attitudes? Nonverbal behavior. McConnell and Leibold (2001) showed that the degree of private prejudice toward Black people held by a White participant predicted less smiling, less speaking, and more speech errors with a Black versus White interaction partner. This pattern of findings tells us that people will often be saying one thing, but sending signals that betray that their true feelings are the exact opposite. How does this impact the interaction, when one's interaction partner picks up on these signals and essentially knows you are untrustworthy?

Dovidio et al. (2002) explored such interactions of people whose overt actions and subtle actions were misaligned in this way. Is this noticed? Which behaviors (the overt or the subtle) have greater weight in one's judgment of the interaction partner? How do people feel in such interactions? To explore this, Dovidio et al. placed two research participants in an interaction where a noncontroversial topic (unrelated to race) was discussed. Of special interest were the dyads that contained one White person and one Black person. They found that for the White people in these interactions, their focus was mostly on their own overt behavior. If they said friendly things, they judged the interaction as having gone well. However, what happens when despite their overtly positive behavior, they have implicitly negative feelings being expressed by their nonverbal behavior? They do not see these reactions. As Dovidio et al. put it:

> Whites have full access to their explicit attitudes and are able to monitor and control their more overt and deliberative behaviors. They do not have such full access to their implicit attitudes. . . . We expect that as a consequence, Whites' beliefs about how they are behaving . . . are based primarily on their explicit attitudes and their more overt behaviors, such as the verbal content of their interaction. (p. 63)

Although the White people do not think they are sending negative nonverbal signals, their Black interaction partners detect them. The interaction partners see both communication channels—what is being overtly said and what is being communicated nonverbally. The results showed that measures of the White person's "hidden prejudice" (and once again, in Chapter 4 we discuss in detail how to measure this) predicted greater amounts of negative nonverbal behavior. And more importantly, the negative behavior that went undetected by the White person displaying it, was easily detected by their interaction partner. The more implicit prejudice held by the White person, the greater the amount of negative nonverbal behavior detected by their Black partner.

When such a misalignment in the verbal and nonverbal behaviors of one's interaction partner is detected, what consequence does this have? Research has shown that a misalignment of one's verbal and nonverbal communication can create feelings of communication awkwardness and general distrust in one's interaction partner. If members of minority groups detect a majority group member saying nonprejudiced things, but have this mismatched to negative nonverbal acts, this creates a sense of distrust, a feeling that a deception is being perpetrated. This reduction in trust from the misalignment makes the

minority group member less satisfied with the interaction than the majority group member, who fails to detect this discrepancy (e.g., Shelton, 2000). Consider the repercussions in an important domain such as medicine. Dovidio and Fiske (2012) argue, "a mismatch between a physician's positive verbal behavior (as a function of conscious egalitarian values) and negative nonverbal behavior . . . is likely to make a physician seem especially untrustworthy and duplicitous to those who are vigilant for cues of bias" (p. 949).

Other research shows that such negative nonverbal behaviors are exaggerated when White participants are also faced with social desirability concerns and are explicitly worried about appearing racist (e.g., Amodio, Harmon-Jones, & Devine, 2003; Goff, Steele, & Davies, 2008). Ironically, one's concerns about appearing biased make one send nonverbal cues suggesting just that, and make one seem less trustworthy, as if one is suppressing true feelings and lying. Evidence for this is provided by Blair and colleagues (2013), who report that Black patients saw White physicians as being less trustworthy and less skilled as a function of the physician's social desirability concerns. The greater the physician's need to hide feelings of discomfort, the more the physician's body communicated that discomfort, and the less satisfied the Black patients were with the medical experience (see also Penner et al., 2010).

Finally, if detecting a mismatch between a person's feelings of discomfort held at the private level and the positive things the person is saying causes feelings of distrust and dissatisfaction, it would make sense if such feelings are then reciprocated. Being the recipient of repeated negative treatment, even if it is nonverbal, can cause ethnic minorities to compensate for this treatment by sending their own signals that they detect the discomfort and feel awkward in such a situation. They may start to signal that they wish to escape this situation in which they are being rejected by their partner (e.g., Shelton, Richeson, & Salvatore, 2005). With one person signaling dislike and the other signaling a desire to escape, such interactions can be anxiety provoking and tense, and can spiral into a negativity that leads both sides wishing to exit. This in turn can strengthen beliefs that members of other groups are less interested in legitimate and equal contact and make people more avoidant of it in the future (e.g., Shelton & Richeson, 2006). In essence, negative treatment by one partner in the interaction will draw out more cautious, uncomfortable, and awkward reciprocal responses from their partner. In one classic experiment, Word, Zanna, and Cooper (1974) found that White research participants displayed a set of negative nonverbal behaviors that communicated discomfort and dislike when interviewing a Black job candidate: an increased rate of blinking, more speech errors, decreased eye contact, and body posture such as leaning away from the person. While they did not detect their own prejudice, it led to disparate treatment and microaggressions. **Microaggressions** are subtle acts, potentially not consciously initiated, that are perceived as hostile and derogatory by the person toward whom those acts are directed (e.g., Sue, 2010). Importantly, Word et al. further showed that being treated in this way produces negative reciprocal behavior. In a second experiment they examined what happens when a person in an interview is treated in this manner and is the recipient of the type of nonverbal behavior these Black interviewees experienced. They found that when you treat people differentially with nonverbal cues, they react in kind. Participants who were the targets of microaggressions were then seen as reciprocating and were judged more negatively than people not treated this way. Negative views led to treating people with negative nonverbal acts (e.g., Dovidio et al., 2002) and elicited negative behavior that confirmed the initial negative view.

Microaggressions: Subtle acts, potentially not consciously initiated, that are perceived as hostile and derogatory by the person toward whom those acts are directed.

The Motivation to Control Prejudice

Any discussion about stereotyping and prejudice is a personal matter, and we can imagine that an individual might not wish to self-disclose this information, and might lie if asked about it. Thus, separate from how any of us think, feel, and value are our social desirability concerns. These are matters of impression management and projecting an acceptable image to others. Will other people perceive us to be biased? Are we going to be rejected for being immoral or "canceled" for failing to meet a cultural ideal about bias? We can imagine that an individual might lie about these matters when asked about it directly. When a person declares "I am the least racist person you will ever meet," there are two distinct reasons they might espouse this concern for prejudice control and a goal of being egalitarian. First, they may be reporting a true belief. Second, the declaration may be strategically false, a deception reflecting simply their concern with *how they look*. A person might be concerned with both of these reasons, or one but not the other.

Plant and Devine (1998) introduced this distinction as two reasons people might be motivated to control prejudice. One is an **internal motivation to control prejudice** (IMCP), which reflects privately held goals to be fair and unbiased, as well as personal beliefs relating to concerns for equality, social justice, and nurturing diversity in one's social life. Members of *other* groups afford one an opportunity to pursue these goals and support those beliefs and are not a threat to one's IMCP, but an affordance. A second is an **external motivation to control prejudice** (EMCP), which reflects a concern with social desirability: concern with doing the socially incorrect thing, worry about the opinions of others, and not wanting to seem biased. Members of *other* groups afford an opportunity for one's social incompetency and bias to be revealed, and create heightened arousal; heightened self-consciousness; and a desire to report beliefs, attitudes, and behaviors that highlight how one is not biased (e.g., Amodio et al., 2003; Bean et al., 2012; Blascovich, Mendes, Hunter, Lickel, & Kowai-Bell, 2001; Mendes, Blascovich, Hunter, Lickel, & Jost, 2007; Moskowitz, Olcaysoy Okten, & Gooch, 2017; Plant & Devine, 2003; Richeson & Trawalter, 2008). Both motives capture so-called principled opposition to bias, but the differences emerge in application of those principles, as seen in work on "laissez-faire racism" and ideological principles (e.g., Bobo, Kluegel, & Smith, 1997; Feldman & Huddy, 2018; Reyna, Henry, Korfmacher, & Tucker, 2006; Sears & Henry, 2003).

Internal motivation to control prejudice: A personally held goal to be egalitarian and fair, with the desire to be nonprejudiced stemming from one's value system and individual needs.

External motivation to control prejudice: A goal to appear egalitarian and fair in the eyes of others and not be seen as doing anything socially incorrect. A desire to report to others that one holds beliefs, attitudes, and behaviors that highlight that one is not prejudiced in order to avoid the anxiety of being labeled biased.

Despite IMCP and EMCP each being motives aimed at controlling prejudice, they do not always produce the same response. IMCP may lead one to say they are an ally, support minority causes, and wish to work toward social justice. The same person whose IMCP leads them to say such things can also have an EMCP that sends the opposite signal. They can appear nervous, anxious, and uncomfortable around a person from a minority group. Even low-prejudiced people with a very high IMCP can have concern with *appearing* prejudiced. In fact, the anxiety associated with EMCP is especially acute for a person who is also high in IMCP. Being nonprejudiced is central to their identity, and having core elements of that identity challenged is especially threatening (e.g., Vitriol & Moskowitz, 2021). Thus, a person high in both IMCP and EMCP will have beliefs and attitudes that denounce bias, but will be exceptionally nervous and anxious about appearing biased and having others

questioning their moral credentials. Despite truly believing they despise bias, the signals sent nonverbally communicate anxiety and bias.

Cross-race interactions can be stressful and arousing for this very reason (e.g., Richeson & Shelton, 2007). Both the majority and minority group representatives in the interaction may have no personal bias toward the other group. However, for the minority representative in this interaction there can be a concern with being stereotyped and treated in an unfair way, and this causes anxiety. Their anxiety may lead them to monitor their behavior to make sure they are not sending any signals that might affirm the stereotypes others hold (stereotype threat concerns; e.g., Steele, 1997) and to monitor their partner's behavior for signs that they are biased. They proactively attempt to make sure the interaction goes smoothly, creating a need to monitor and regulate behavior that does not exist in same-race interactions (e.g., Richeson & Shelton, 2003; Shelton et al., 2005; Shelton & Richeson, 2006; Taylor, Garcia, Shelton, & Yantis, 2018; Trawalter & Richeson, 2006; Vorauer, Main, & O'Connell, 1998). This is hard work. For the majority-group representatives there is social anxiety introduced by the interaction associated with potentially being seen as prejudiced, an anxiety that does not exist in same-race interactions. Their anxiety may lead them to engage in impression management to make sure they say only things that can lead to a positive impression and give off no whiff of racism. This is hard work. And may mismatch what they are communicating nonverbally. Such interactions are complex.

Richeson and Shelton (2007) show that all this hard work is draining and actually makes people less effective at performing other tasks that require self-regulation. In one experiment it was illustrated how a cross-race interaction led to decreased performance on a subsequent task that required cognitive control (the Stroop task, reviewed later in this chapter), thus indicating a reduced capacity to self-regulate among people who needed to regulate their stereotypes during the interaction. The devotion of cognitive resources to monitoring the expression of stereotypes and disproving that one may be biased during the interaction worsened performance on a task assessing executive functioning. As Richeson and Shelton stated:

> Engagement in one task that requires self-regulation (e.g., inhibiting behaviors, thoughts) impairs later tasks tapping the same resource. Self-control draws on a central executive attentional resource that can be depleted. Based on the model, therefore, interracial contact impairs performance on tasks that require executive control because individuals engage in self-control during the interaction, which depletes their executive attentional capacity. (p. 317)

They also show that such interactions are distracting, with attention diverted to monitoring these secondary matters of impression management (see also Vorauer & Kumhyr, 2001). Concerns with social desirability from both interaction partners—not affirming a negative stereotype, not seeming to endorse a negative stereotype—lead the interaction to be fraught. Interestingly, Richeson and Shelton report that Black research participants actually preferred a more openly (as compared to less openly) biased White person as an interaction partner because they knew where the person stood. There was no misalignment of verbal and nonverbal behavior, and no feeling of distrust and deception. No need to regulate. At least they knew why the interaction was negative.

Avoidance

The anxiety that social desirability concerns create in cross-race interactions has so far been said to have consequences such as deception, creating distrust, triggering arousal, draining

people, and distracting them. As a result of these unpleasant consequences an individual will, often unintentionally, avoid people who are not from their own group. Research in a wide variety of domains (e.g., avoiding eye contact, sitting farther away, ending conversations sooner, exiting an interaction more quickly, avoiding an encounter altogether, speaking less) shows people engaging in such avoidance. At times the avoidance is explicit, as in a classic study of attitudes from the 1930s where hotel owners expressed a desire to avoid contact with minorities by denying them a room at the hotel. This type of denial and avoidance can occur even when the person running the hotel is not overtly prejudiced (e.g., Howerton, Meltzer, & Olson, 2012). At other times people can be subtly avoidant. For example, one way a person can be avoidant is by exiting a situation where social anxiety is present. Peck and Denney (2012) provided evidence of such avoidant behaviors in the health care professions. They studied doctor–patient interactions to see whether doctors exited cross-race interactions more quickly than same-race encounters. To assess this Peck and Denney looked at the medical interviews conducted within the doctor's office. This is not arbitrary chitchat but follows a structure and format in which the doctor has been trained. It should *not* vary based on race, yet it does. Looking at 221 doctor–patient encounters, they found that the amount of patient input solicited and the amount of control exerted by doctors in the interaction fluctuated as a function of race. This resulted in non-White patients having shorter medical visits than White patients (23.9 vs. 28.5 minutes). The 4.5-minute difference represents a 20% shorter visit for patients who are not White (two-thirds of whom were Black).

A similar type of subtle avoidance can emerge in job interviews that invoke social desirability concerns. Hebl, Foster, Mannix, and Dovidio (2002) showed bias in a job interview toward gay and lesbian job applicants, despite participants not realizing they had bias. The bias did not manifest in the form of overt hostility and dislike, but in the form of discomfort and avoidance. Less interest was shown in gay applicants; interactions ended sooner. Specifically, they found that fewer words were spoken by a heterosexual interviewer to gay than straight applicants. Also, the length of the interview was shorter with gay versus straight applicants (M = 245 seconds vs. 383 seconds). The interviewer exits the interview sooner when they are with a member of a social group that is not their own. Avoidant strategies are also indicated by measures showing that the potential employer exhibits less eye contact, and acts more standoffish, with a gay applicant. While clearly prejudice behavior such as lying about the availability of the job, or not being allowed to fill out an application when asked, is not seen, subtle negative behaviors are.

Avoidance is also seen in an experiment by Kawakami, Dunn, Karmali, and Dovidio (2009). Their participants met two confederates posing as other participants . One confederate was Black, the other White. The situation was scripted so that the Black "participant" needed to leave the room momentarily and on the way out gently bumped the White confederate accidentally. Some participants saw only this accidental bump and then were asked to choose who they would want to be their partner on the next task. Half chose the Black confederate and half chose the White confederate. Other participants saw the same accidental bump but also saw the White confederate *utter a racist comment about the bump while the Black confederate was out of the room*. Some heard a moderate racial slur ("I hate it when Black people do that") and some heard an extreme racial slur (the use of a word widely regarded as extremely offensive in the English language). Who does the research participant pick to be their partner? Rather than rejecting them, participants embrace the person who made the racist remark. In both slur conditions they avoid the Black person and choose the racist White person (63% of the time). Observing a racist act makes racism salient, and highlights the racial nature of the interactions that can follow. That anxiety and threat makes people avoidant. So much so that they choose to partner with the racist and avoid the target of the

racism. Of course, when asked to predict how they would act in such a situation, people say they would summarily reject the racist. Similar avoidant behavior is seen in an experiment by Plant and Devine (2003) where research participants were asked to return to the lab at a later date for an interaction. They found that as participant anxiety about an upcoming interracial interaction increased, they were less likely to return for the interaction.

Finally, avoidance can also be seen in very low-level responses. In one experiment, Bean et al. (2012) found that White participants who were high in EMCP had a bias in visual attention to faces of Black men that was indicative of threat. Using eye tracking they revealed that, without realizing it, participants would shift their gaze toward faces of Black men and away from faces of White men within the first three-quarters of a second from when the faces appeared. However, at about 1 second (as conscious control began to set in) their gaze quickly shifted so that they became avoidant of the faces of Black men. The immediate response was as if a threat was present (focus attention on the threat), and the more controlled response was avoidance (to divert one's gaze).

Cross-race interactions become an arousing chance for bias and social incompetency to be discovered (e.g., Plant & Devine, 2003; Vorauer & Kumhyr, 2001). Research has established that although external motivation to avoid prejudice is not a general form of arousal that cuts across domains, it is predictive of arousal in intergroup responses, even when White participants are simply briefly shown faces of Black men (e.g., Amodio et al., 2003; Bean et al., 2012; Plant & Devine, 2003). This type of arousal makes individuals motivated to avoid experiencing the anxiety—they avoid cross-race interactions. In this way, people who report themselves to be high in EMCP can be quite poor at controlling the bias they so desperately want to control. Ironically, the very social desirability concerns that they report as making them legitimately not want to be biased will make them anxious and threatened. These feelings can make them avoidant, which is a type of bias. Even when not avoidant, it can make them signal unease and discomfort to an interaction partner, another form of bias. Thus, their self-report is unreliable and inaccurate; it is called into question by their nonverbal behaviors. This is a type of deception. People say they are not biased, they intend to be not biased, but they act in a biased way anyway. However, it is not a deliberate deception. They did not intend to lie.

AVERSIVE RACISM AND SELF-DECEPTION

In the previous section people were shown to lie to others due to social desirability concerns. At times the lies are deliberate (as with the Bradley effect, or minority influence, or conforming to offensive jokes or politically correct attitudes). At times the lies (e.g., saying all is good while displaying nonverbal signaling of dislike, avoiding others) are an outgrowth of unintended processes, such as when private feelings misalign with public norms and cause anxiety. In all of those examples the lie was that a person's true feelings were being concealed from others. Here we turn to lies to the self: where people deny their true feelings in both what they self-report to others and in what they consciously admit in their private thoughts.

Self-Deception and the Desire to Look Good

People at times cannot handle the truth. They deny and suppress inconvenient truths. *People often lie to themselves*, which means they do not recognize their act of self-deception. The lie is engaged to protect self-esteem from damage. Above you were asked to imagine a person

without bias, but who had anxiety about appearing biased. Here, we ask you to imagine a more complicated person. Imagine a person who legitimately does not want to be biased and has anxiety about being biased, but who nonetheless has bias that they do not recognize. This is a person who wants to be fair and unbiased and without prejudice, yet deep down they have such biases and prejudices that they do not consciously see. These unwanted thoughts are hidden from the self because they violate important self-standards and values.

Self-deception can shield one from unwanted and non-normative beliefs and feelings. People who desire to be low in prejudice often do not wish to face the reality that they have biases and can at times respond in biased ways. To protect themselves from this negative view of the self, they self-deceive by believing wholeheartedly in their nonbiased sense of self. When you do not want to face your own biases and so you repress and deny them, you have engaged in what is called **aversive racism** (Dovidio & Gaertner, 1986; Gaertner & Dovidio, 1986). One is not only unable to see one's biases, but what one does see is that one is not prejudiced and a need to reject prejudice. To do this people will often (1) exaggerate their own sense of self as unprejudiced, and (2) believe that bias in the culture is extreme by pointing to terrible exemplars (so one's own views are positive in comparison). When engaged in this type of self-deception, the suggestion that one is prejudiced will be experienced as highly aversive, yet at the same time one does in fact possess unconscious thoughts and feelings that label another group as aversive. Despite one not consciously seeing it, one holds negative outgroup beliefs and feelings and consciously has strong egalitarian beliefs.

Aversive racism: Implicit prejudice that emerges among people for whom the suggestion that they are prejudiced is aversive, who simultaneously have unconscious thoughts/feelings that another group is negative or to be avoided. It is a dissociation from the reality that biases exist in the self and reflects a legitimate desire to lack biased tendencies.

As just described, overtly the aversive racist is convinced of their lack of bias and has no overt antipathy. Instead they experience discomfort and anxiety. As discussed earlier, when people have anxiety due to consciously worrying about being seen as prejudiced, they can be avoidant, and they can feel threatened and anxious. Aversive racists experience these same subtle avoidant responses, but for different reasons. Rather than the anxiety arising from a bias they know they have and do not want others to discover, the anxiety arises from a conflict they cannot see. Aversive racism describes a "new" type of bias that complements old-fashioned racism. This is not to argue that the more prototypical forms of explicit and overt prejudice no longer exist. Certainly, prejudice can be expressed openly, and the willingness to do so waxes and wanes with the times (and we are unfortunately in a time where it seems to be waxing).

Aversive racists are continuously providing evidence to support their self-deception that they reject prejudice. When they find themselves in situations where discrimination would be obvious and the social norms to reject bias are strong, aversive racists will act in clearly nonracist ways. They are motivated to illustrate their egalitarian intent. They can point to nonracist beliefs, actions, and attitudes in these situations as a way to avoid facing the bias lingering below the surface. Pearson, Dovidio, and Gaertner (2009) argue that what the aversive racist does not see are the subtle ways in which bias can often leak out through how one acts. In situations that are clearly ones where bias can manifest, they act in an unbiased way that can be contrasted with the bias seen in others. However, there are other situations, that are less clearly about race, where the aversive racist does not think they have acted in a biased way, but an experimenter can easily observe it. Pearson et al. reviewed four types of situations in which aversive racists produce discriminatory responses without realizing their bias. Three situations in which bias can leak out without one realizing it are "situations in which normative structure is weak, when the guidelines for appropriate behavior

are unclear, [and] when the basis for social judgment is vague" (p. 318). A final way bias leaks out is in a situation in which one's actions can be justified or rationalized on the basis of some factor other than race. We begin our review of how to detect aversive racism there.

Bias Can Leak Out When It Is Unclear That the Situation Is about Race

When a situation is clearly about race, an aversive racist knows to control how they act. When the situation is not seemingly about race, they are less controlled. What is meant when we say "the situation is not seemingly about race"? It is when race is confounded with other possible reasons to disagree with or dislike a person. When people are **using legitimate issues that are ambiguous to mask bias**, pointing to their principled stance as the reason for the dislike. For example, it is unclear the situation is about race when there are reasons other than race present for disparaging a person—their political views, their inexperience, their having been accused of a crime, their acting badly at an awards show. These are all legitimate reasons to dislike a person that do not need to invoke race. With "cover" being provided by such legitimate reasons, race can be allowed to drive how one responds without fear of being labeled racist. However, such responses are indeed racist if one responds differently when a Black person performs the behavior in comparison to a White person. If your response to a Black politician seeking office, or a Black academic receiving an award, is "they lack experience, I cannot support it," you would be exhibiting aversive racism if your response to a White politician or academic with the same inexperience was "what a young superstar." In such examples the response of the White person to the Black person's situation seems to that White person to not be about race, allowing them to not see that the response is a biased one. Such situations where race is a possible reason for dislike, but another cause is also seemingly legitimate, allow us to diagnose aversive racism since the aversive racist will not see (or control) their bias. They will just lean on the seemingly legitimate explanation.

Using legitimate issues that are ambiguous to mask bias: Pointing to existing facts as justification for a response, when in reality that fact is used as a mask to hide a bias that is the true basis for the response. For example, saying existing policies are the reason one dislikes a candidate when it really is gender bias.

It is not that one dislikes Black political candidates but this candidate's set of policy positions. It is not that one dislikes Black men but one sees the data as showing that Black men commit more crime. It is not that one disfavors Black applicants for college but the qualifications show the White applicants to be superior. It is not that one does not want more Black people hired at work but that one has a principled opposition to the policy of affirmative action since it is not based on merit. Or to make things very real as I write on a Sunday morning in March of 2022, it is not that one prefers White Ukranian refugees over Syrian or Sudanese refugees but the conditions of the Ukranians warrant special treatment. Of course, it is possible that any of the above expressed beliefs are not race based and are formulated based on the data and good evidence. One could truly oppose affirmative action because of a principled feeling about meritocracy that has nothing to do with race. Not all people espousing such beliefs above are racist. However, aversive racism research is able to show *when* it is race, since the evidence used to justify the choice is shown to shift to whichever available information is negative about the stigmatized group. For example, if a White person opposes affirmative action because it is not meritocratic, yet supports other types of deviations from meritocracy when they benefit White people, then it is race not meritocracy that is the real issue, and meritocracy is used as diversion. One might oppose a mayoral candidate who is Black using their positions on issues as a justification for one's opposition to them, yet if a White candidate with very similar positions on the same issues

receives one's support and is deemed acceptable, then it is race not policy that is really directing one's response. The seemingly legitimate cause is shown to be a crutch. McConahay and Hough (1976) stated:

> Behaviorally, it is a set of acts (voting against black candidates, opposing affirmative action programs, opposing desegregation in housing and education) that are justified (or rationalized) on a nonracial basis but that operate to maintain the racial status quo with its attendant discrimination against the welfare, status, and symbolic needs of blacks. (p. 24)

To illustrate this phenomenon, McConahay (1982) performed an experiment that asked White research participants about the policy of using buses to more equally distribute children across schools. In 1970s Louisville, Kentucky, where the study was being conducted, this meant the percentage of Black children in predominantly White schools was increasing due to busing. There are many issues relating to one's own self-interest that one could say is a reason to oppose such busing that are not racially motivated. McConahay examined issues such as having school-age children, being the parent of a child who would be bused, having an interest in maintaining the social stability of the neighborhood, owning a home in the neighborhood, and so on. However, White participants who claimed their opposition was about the policy and not about race were shown to have it truly be about race. McConahay stated:

> Various measures of high and low self-interest among whites were virtually useless in discriminating degrees of support or opposition. On the other hand, measures of racial attitudes were correlated strongly and consistently with the anti-busing position: the more racist, the more opposed. In short, it is not the buses, but the blacks that arouse the ire. (p. 714)

In this instance the White participants truly believed they opposed the practice of school busing, and that objections to this policy had nothing to do with race or the increase of the percentage of Black children in their own child's school. However, opposition to a legitimate issue—busing—was used to mask an illegitimate bias against a group of minority children from entering one's circle. The participants rationalized their preferences on the basis of political beliefs rather than race. A White person will allow racist feelings to seep out if they think they are expressing beliefs regarding issues they do not see as about race. A similar illustration using the more modern example of opposition to Obamacare was provided by Knowles, Lowery, and Schaumberg (2010).

In a second example, Hodson, Dovidio, and Gaertner (2002; see also Dovidio & Gaertner, 2000) showed how people will justify a biased college admission decision by pointing to the credentials of the candidate as the reason for the applicant's rejection. From an aversive racism framework, we would predict that when the qualifications of the prospective applicants *were clear*, *an unbiased* decision would be made—that is, if an applicant had excellent credentials on every dimension (e.g., both grade point avaerage [GPA] and SAT scores were excellent), there would be no bias. It would only be when a résumé was *questionable* that decisions might reflect bias due to the ability to point toward the insufficient qualifications as the justification for the rejection. Such studies allow us to see bias in action because the White participants show a tendency to reject the Black applicants in favor of White applicants regardless of why the Black applicants lack strength. If a Black applicant has stronger SAT scores but a weaker GPA than a White applicant, White participants say GPA matters and choose the White applicant. Yet, if a White applicant has the exact same qualifications of stronger SAT scores and a weaker GPA than a Black applicant, the White participants

now say SAT scores matter and again prefer the White applicant. A pro-White choice is made no matter the data but justified as purely data driven.

In a third example, this has also been shown in hiring decisions. Son Hing, Chung-Yan, Hamilton, and Zanna (2008) investigated discrimination against Asian job applicants in Canada and found that when assessing candidates with identical qualifications, evaluators recommended White candidates more strongly for the position than Asian candidates with identical credentials. Yet they claimed it was credentials that were the basis for the decision. Even though the reality was that whenever the credentials on which the job applicant was best shifted, the participants shifted what credentials they chose to highlight as the basis for the decision. If a White person was superior to the Asian candidate on quality "*x*" but not "*y*," people justified the choice of the White candidate by saying it is clearly quality "*x*" that matters most and is therefore why I selected the White candidate. However, when the White candidate was superior on quality "*y*" then the importance of quality "*y*" was now seen as self-evident and the correct basis for the choice. The same credentials disparaged when held by an Asian applicant were seen as strong when held by a White applicant. The decision always seemed to the participant to be based on evidence, but they could not see how their assessment of the evidence was biased by prejudice. This bias was greatest among participants who scored highest on measures of implicit prejudice.

Self-Deception Can Leak Out When Norms of How to Act Are Unclear

Another time aversive racism can leak out is when there is a **weak normative structure**. Situations vary in regard to how clearly there are rules for how to act, with some situations constraining us entirely, and others where there are no rules (and all points in between). The normative structure is said to be weak when the rules for how to act are poorly defined or nonexistent. For example, if you are alone and come across a person in dire need of help, the norms tell us it is incumbent on us to call for help (at a minimum). If you are one of thousands of people at a festival and see the same person in dire need, then norms of how to act are far less clear. Aversive racism can be diagnosed in such situations, where there is uncertainty about how to act. Helping situations provide a nice way to illustrate the point: Does racism leak out by failing to help a Black victim when one would help a White victim?

Weak normative structure: Situations vary in regards to how clearly rules specify how to act. Some situations' rules constrain us entirely, others situations have no rules. The normative structure is said to be weak when such rules are poorly defined or nonexistent (such as emergency situations with many other bystanders also present).

For example, when a person is in a situation where norms of helping are clear, it would be inappropriate to deny help, and failure to do so when the person "in need" is a member of a minority group would clearly suggest the inaction was a form of bias. However, if your role as a bystander less clearly dictated intervention was required (e.g., Darley & Latane, 1970; Latane & Darley, 1970), then failure to act based on race can be masked by the weakness of the situational norms. Gaertner and Dovidio (1977) provided one of the first experimental illustrations of aversive racism using such a situation. Research participants were White female undergraduates who overheard a supposed emergency in which several chairs seemed to fall on either a White or a Black female confederate. The participants were either alone or in the presence of two other bystanders. No differences in helping behavior were found when alone: The norms were clear. This is a helping situation, and when alone you must be the one to help. When others are present the norms about helping are weaker, and a helping situation that was not seemingly about race can now be seen to be clearly about race. A Black person in need of help in such a situation was provided that help about half as often as a White person.

Gaertner (1973) provided another example. Research participants were U.S. citizens who were called on the telephone at their home with a wrong number that presented a crisis situation to the call recipient. The person calling (long before the days of cellular phones) claimed to have a car broken down on the side of the road and had walked to a phone booth and used their last coins to make this call to what they thought was a service station. Having dialed the wrong number, and now out of coins, they had a simple request: The participant was asked to call the service station for them and report the incident so that help could be sent. The phone number for the "service station" was actually the research lab, where it could be recorded how many people called. Here are two additional keys to the research: (1) half of the participants heard callers with voices that indicated they were likely Black men and half of them heard voices that indicated the person was likely a White man, (2) half of the people called were White and registered as political liberals and half were White and registered as political conservatives. Gaertner found two distinct types of racism evidenced by these unsuspecting research participants. One type was to simply hang up on the Black caller before the emergency situation could be explained. Another type was to listen to the complaint, but then to not take action if the caller was believed to be Black. Conservatives did not differ in hang-up rates, but were far less likely to help a Black caller who explained they needed help (65%) than a White caller (92%). This is an example of more overt bias. They heard that help was needed yet did not offer it at the same rate for Black versus White people. Liberals, in contrast, were just as likely to help each group if they waited to hear the problem—however, they were far more likely to avoid hearing the problem if the caller was Black—they hung up early on a Black caller (20%) more often than a White one (3%). They may believe the inaction was caused by the call being a scam, not because the person was Black. But in this normatively weak condition, the pattern of behavior to these callers suggests a subtle bias is present.

Aversive Racism Leaks Out When the Basis for Social Judgment Is Vague

Aversive racism also manifests when the stimulus being observed is ambiguous. For example, Dovidio and Gaertner (2000) showed research participants the profiles of the personnel at a peer counseling center that were supposedly culled from their job interview. In the key conditions of this experiment the qualifications of the person revealed in their interview showed them to be ambiguous in regard to how well-suited they were for the job. Some of these were White and others Black job candidates. The results were clear—though participants reported having no bias, there was in fact a pro-White bias. Ambiguously qualified Black candidates were evaluated more poorly and were less likely to be recommended for the job than White candidates with the same qualifications. When the records were not ambiguous, and the applicants were clearly strong, there was no bias.

AUTOMATIC PROCESSES AND THE NOTION OF "BELOW-THE-SURFACE" THINKING

The cognition that we experience consciously (such as experiencing the sky as blue) is the end product of a series of processing steps that deliver that experience. A focus on the end result, the conscious experience, can obscure the fact that the experience arises from cognition that we (1) do not see, or (2) cannot see. As described in Chapter 1, there is a ***phenomenal immediacy*** that makes the conscious experience seem to just appear suddenly, as if delivered by the properties of the stimulus and not produced by the processes of the mind. This does

not mean people do not ever recognize the role of their own perception and attention and learning processes in producing what they consciously experience. However, when they do, their assumptions about how they arrived at their conscious experience are often incorrect. It is difficult to introspect upon such inaccessible processes, so people lack the ability to see and to know what to report about how those processes unfold. Yet they still *feel confident* that they know. Humans have a lifetime of experience convincing themselves that they know how they think and what they think. This "feeling of knowing" about their cognitive process gives people high levels of confidence in their conscious experience. The beliefs people hold about their cognitive processes are often wrong, despite the confidence with which they are held.

Nisbett and Wilson (1977; see also Wilson & Brekke, 1994) provided an important illustration of just how much people lack awareness of the processes contributing to their perception of one another. In one clever experiment, Nisbett and Wilson asked participants to memorize a series of word pairs. For some of the people the pairs of words included the pair "ocean–moon." Shortly after finishing this task the participants were asked to answer some mundane questions, and this included among them a question that asked them to name a laundry detergent. The number of people who responded with the brand "Tide" was double for people who had previously seen the word pair "ocean–moon." Are people aware that the word association task influenced their responses about laundry detergent? No. They have a wholly different explanation for the cognitive process that brought that brand to mind. Nisbett and Wilson found that people not only lack awareness of an influence on them (such as thoughts of the ocean or moon unknowingly influencing production of the brand "Tide") but they are also unable to accurately report on the nature of the influence when they correctly suspect one exists. In one experiment students watched a video of a teacher who spoke with an accent. Some people saw the teacher respond in a "warm" way, others saw the teacher respond in a "cold" fashion. The warm/cold behavior determined how much they liked the teacher, and the degree to which they liked the teacher influenced other ratings, even ratings of the teacher on dimensions that should not have been influenced—his attractiveness, how much they liked his accent. The important point is not the fact that such *halo effects* exist, where ratings on one dimension spread to other dimensions. It is that participants cannot accurately guess the direction of the influence! They know their ratings of the teacher are biased, but they believe it is the teacher's attractiveness that influences liking, when it is actually the other way around. When asked to reflect about what we do when forming impressions, we do not have good access to what it is we are doing, and cannot reproduce it accurately when asked.

In the above examples people lack access to accurate knowledge about their cognition because they do not comprehend the actual mechanisms involved. However, a separate reason people lack access to accurately describing their cognitive processes is because they are unable to recognize that any mental activity is taking place. We turn now to a focused discussion of cases of cognition being inaccessible due to its invisibility—to cognition that is automatic. An ***automatic process*** is not simply a cognitive process for which one lacks conscious awareness. Since most cognitive processes contain some components that occur outside of awareness, practically all processes would be called automatic if automaticity was defined by a lack of conscious awareness at any point in the process. For example, even complex behaviors such as driving lack awareness of one's responding at times. If such actions were to be called "automatic," it would render the term so vague as to be useless. A cognitive process must have all four of the features described below to be defined as automatic (Bargh, 1984, 1989, 1994, 1997). First, an automatic process is one that lacks conscious intent in that it is triggered immediately and directly from stimuli in the environment

rather than initiated by a conscious choice. Second, automaticity is marked by a lack of control—once triggered the process will run to completion without disruption (even if one wanted to control it). Third, automatic processes are efficient in that they cannot be disrupted by other ongoing mental activity and usurp very little processing effort. Finally, they occur without awareness; consciousness is not involved at any stage of processing. Driving is not an automatic process because, despite at times lacking conscious monitoring or attention, it is consciously willed (the mere presence of a car does not cause one to jump in and begin driving) and consciously terminated.

The root of the concept of an automatic process can be traced to Charles Darwin's (1872/1998) description of the nonverbal emotional signals humans send to communicate with important others. When describing how complex behavior such as this is routinized, Darwin stated, "some actions, which were at first performed consciously, have become through habit and association converted into reflex actions, and are now so firmly fixed and inherited, that they are performed, even when not of the least use" (p. 45). Similar language is used by James (1890/1950) to define habit: "a strictly voluntary act has to be guided by idea, perception, and volition, throughout its whole course. In habitual action, mere sensation is a sufficient guide, and the upper regions of the brain and mind are set comparatively free" (pp. 115–116). Wegner and Bargh (1998) defined an automatic process as a *mental habit*, "patterns [that] become the deep grooves into which behavior falls when not consciously attended" (p. 459). How does a process become automatic? Through practice, repetition, and habit. Bargh (1990) proposed that a response that is routinely paired with a specific set of environmental features can, over time and practice, lead to the activation of the response given the presence of those environmental features. Let us next turn to examining in detail each of the four elements described above as necessary for identifying an automatic process: lack of conscious intent, efficiency, lack of control, and lack of awareness.

Lack of Conscious Intent

An automatic process is not consciously initiated, but triggered by the stimuli to which one is exposed. When one looks up at the sky one does not consciously intend to perceive its color. In our perception of the people we encounter there is also a good deal of processing that proceeds without our conscious intent: assessing race, gender, facial expressions, and so on. What about complex inferences such as beliefs about a person or attitudes toward a social group? Can these form without intent? How can we illustrate that a process is initiated without intent?

Let us start with the process of detecting a person's facial features. One way that researchers illustrate that facial features can be detected without conscious intent is to present an image of a face subliminally—below the person's threshold for conscious recognition. This is called **subliminal presentation**. If one never consciously detects the presence of a stimulus, yet that stimulus is able to trigger responses associated with it, then the processes associated with detecting it and responding to it must not have been consciously intended. For example, Murphy and Zajonc (1993) subliminally showed faces to research participants, manipulating whether the facial expressions depicted positive or negative emotions. Since the faces were not consciously seen, the participants could not intend to detect the expression and could not intend to have a mood triggered in them by virtue of the expression. Nor could they intend to be influenced by their mood in their judgments of some new and unrelated picture. Nonetheless, the valence

Subliminal presentation: When stimuli appear and disappear so quickly that they are never able to be consciously detected, yet they are detected outside of conscious awareness by the perceptual apparatus.

associated with the faces was shown to influence the judgment of an ambiguous (neutral) stimulus. The stimulus was liked more when preceded by an "unseen" face with positive (as opposed to negative) valence.

Subliminal exposure to stimuli is a useful way to show lack of intent, but not very typical in everyday life. But the lack of conscious intent is easily demonstrated in other ways, most powerfully when people engage in thoughts and actions relating to other people that are the exact opposite of what had been consciously intended—that is, a good way to prove a response is unintended is when people explicitly intend to do something else. Wegner (2002) explains a host of "mystical" phenomena observed through the ages as cases of a person having a conscious intent, and then unconsciously acting against it. The person then misattributes the unexpected behavior to the supernatural rather than to the powers of the automatic processing system. Divining rods, Ouija boards, automatic writing, séance tables lifting or spinning, pendulum divining, water dowsing, and alien hand syndrome all share a common cause. One does not intend to move the object, yet it moves. How can this happen? People imbue the objects with "spiritual" power to explain it. But spirits and magic are not needed. Just because one does not consciously intend to move an object (such as with pendulum swings being used to decide a baby's gender) does not mean that the movement reflects something magical. Instead, people may consciously intend to keep their hand still and let the pendulum "speak," all the while allowing their behavior to be unconsciously guided by their own expectations and desires. One may have an expectation of the result (such as I expect the baby will be a boy) that is never consciously recognized, and such an expectation can cause shifts in muscular movement that produce the expected result (the pendulum swings in the direction that indicates the baby will be a boy). There is no feeling of personal agency or willing. In fact, the agency is to do the exact opposite. People feel as if they know they are not causing the motion (e.g., Ansfield & Wegner, 1996). And because of this **feeling of knowing**, the response feels magical. But it is not. It simply lacks conscious intent.

Feeling of knowing: When one has a conscious intent/belief that one knows why one acts, even when the cause is an automatic process one does not see, and this sense of knowing is wrong. This produces the sense that some outcomes are magical, mystical, or spiritual because one "knows" they were not intended.

There are many examples of people responding in ways that are the exact opposite of what they consciously intend to do. As another interesting example, people often wish to conceal how they truly feel about someone or something. Rather than "wear the heart on the sleeve," a person may want to keep their feelings personal, or even communicate the opposite (acting pleasant when encountering a person who is despised). Concealing emotions when we engage in deception may be what we intend, but we unintentionally continue to send signals that reveal the emotion we wish to conceal. Ekman and Friesen (1969) proposed that while it may be possible to verbally suppress certain ideas or emotions, the body is not always a willing accomplice to our attempts at deceit. Nonverbal signals communicate our true feelings and beliefs with others even when we intend to hide those beliefs through what we say (e.g., DePaulo, Kashy, Kirkendol, Wyer, & Epstein, 1996; DePaulo, Lanier, & Davis, 1983). Despite our intentions, it is difficult to monitor our nonverbal behavior, and the information we intended to keep hidden leaks out. We saw this above when reviewing cross-race interactions. Although we do not intend to send signals, perceivers detect the unintended messages coming from different channels, such as body posture and facial expression (e.g., Ekman & Friesen, 1974; O'Sullivan, Ekman, Friesen, & Scherer, 1985).

It is likely not surprising to learn that we conduct communication with others through a nonverbal language—our personal sign language. Our bodies, faces, social distance, and tone of voice communicate information that interaction partners are constantly sending to and receiving from each other. A wave, a wink, leaning in, a furrowed brow, the middle finger, a shameful smile, pushing away, pulling toward, an outstretched thumb, and a look of

disgust are all part of a silent exchange in social interaction. What may be surprising is *how* silent these exchanges are. One illustration was the nonverbal anxiety communicated during cross-race interaction (e.g., Dovidio et al., 2002; Richeson & Shelton, 2007). Chawla and Krauss (1994) provided an experimental illustration of how people unintentionally send and detect nonverbal cues in a domain not involving race. Research participants were to determine whether a person was delivering a rehearsed speech or speaking spontaneously. Two tapes were created of the same speech, one delivered spontaneously, the other a recreation of the same speech by an actor. Each research participant rated how spontaneous the speech was, and the experimenters then correlated these ratings with the use of nonverbal cues by the speaker. Participants were not intending to use nonverbal cues to help them make this decision, nor were they intending to focus on a specific type of nonverbal cue to help them make this decision, but they did so. Spontaneity ratings made by the perceivers were significantly correlated with certain types of gestures and pauses. These were hand gestures and pauses in speech known to be related to problems with lexical access (trouble pulling words and thoughts from memory). Without intending to, perceivers scanned the behaviors for these cues, and used these cues to help them decide whether the behavior was spontaneous. The types of cues used were nonverbal acts such as observing someone tilting their head when trying to think of just the right word, or someone pausing as if trying to "off-the-cuff" think of a good example.

Efficiency

Efficient processes are able to operate even in the face of limits to one's processing capacity (like being rushed, working on many tasks simultaneously, having divided attention, etc.). An efficient process requires little mental energy and effort and is not constrained by ongoing mental activity; it runs to completion without being disturbed regardless of processing constraints that disable other forms of cognition. Color detection is an efficient process. You can detect that the color of a passing car is red while simultaneously straining memory for the year Martin Luther King was murdered (it was 1968). In person perception research, many of the processes used to characterize other people (and the self) proceed with such efficiency. Being lost in deliberation, or straining to retrieve information from memory, or trying to remember the grocery list does not incapacitate the ability to form impressions of other people (e.g., Uleman, Hon, Roman, & Moskowitz, 1996).

Efficient processes: A process that operates even in the face of limits to one's processing capacity (like being rushed, multitasking, or having divided attention) because it requires relatively little mental energy and effort and is not constrained by ongoing mental activity. It runs without being disturbed regardless of processing constraints or cognitive load.

For example, Bargh and Thein (1985) provided evidence that the processing of highly relevant information is efficient. For some of their research participants' traits related to honesty were highly relevant; other participants had no particular affinity for the trait of being honest. Participants then read a set of behavioral descriptions about other people. A given set had 24 behaviors, 12 of them implying a relevant trait (such as honest), 6 implying an inconsistent trait (dishonest), and 6 were neutral. Descriptions were presented one at a time on the computer. Some people were asked to read the sentences in only 1.5 seconds, making the task highly demanding of their mental energy. Such people were described as being under time pressure to respond, having to make a rapid response. Although rapid response is not exactly the same thing as making responses when inundated with large amounts of information, for the sake of convenience we group all such types of responses that place a person under highly limiting conditions as a response made under ***cognitive load***. For other participants the task was performed while they were not under cognitive load. In

summary, there were people who either did or did not value honesty who received information about another person that was either relevant to honesty or not, and they received this information while either under or not under cognitive load.

If forming an inference is an efficient process, it should occur regardless of whether one is under cognitive load. If not efficient, the time pressure (load) should create limitations that disrupt one's ability to process the behaviors and form an impression. The data show that people who do not value honesty have no trouble processing the information when there is no load (so they come to see the person as honest), but when doing the same task under cognitive load they no longer come to form an impression of the person as honest. Their ability to form a coherent impression is interfered with by the limit placed on their processing. However, people for whom honesty was a relevant trait do not have such an impairment. For these people the proportion of honest traits that were presented was able to be detected and used to guide their impressions. This was true both when cognitive load was absent and when it was present. This suggests that processing information that one deems as highly relevant is efficient. It happens even when responding occurs too rapidly to stop and think deeply.

The **cocktail party effect** is another illustration of the efficiency of self-relevant information. You have likely experienced your own ability to detect your name being spoken in a loud and crowded room when you were engaged in a conversation with someone and not paying attention to what people engaged in other conversations were saying. Yet somehow when they speak your name, it turns out that you were, at some level, attending to what others were saying in the din. How is this possible? Much of perceptual experience takes place prior to your conscious awareness getting involved. The mind efficiently perceives many more things than get reported to consciousness. This ability is linked to a differentiation between short term and iconic memory. Due to the huge amount of information that bombards our senses at any given moment we have developed the ability to store large amounts of information, for very brief periods, without consciousness getting involved. The vast sensory storage house of visual information is known as **iconic memory** (called *echoic memory* for auditory stimuli). Once information enters this storehouse, people are able to "decide" what information, from this bombardment, enters consciousness, capturing our focus of attention, and is represented in short-term memory. These "decisions" about what information to keep and what to filter out occur prior to conscious reflection and are done efficiently, without being constrained by conscious mental activity. Thus, although you may not have been consciously attending to another conversation in the room, the contents of other conversations were being scanned and placed in the iconic storehouse. When that content is self-relevant, perception and attention shift to alter what enters consciousness.

Cocktail party effect: Phenomenon whereby one detects one's name spoken in a loud and crowded room when otherwise fully engaged and not paying attention to what people in other conversations were saying. Yet one's name jumps out from the din. It indicates the vast attentional capacity humans have beyond what they consciously recognize.

Iconic (echoic) memory: A vast sensory storage house of information where perceived stimuli are held before consciously detected. Once in storage people can "decide" what information enters consciousness, capturing our focus of attention, and this is represented in short-term memory. This vast amount of information is in storage only very briefly.

Lack of Control

Lack of control refers to one's inability to stop a cognitive process from happening. Once the stimulus that triggers the process is present, one cannot stop the process from starting. Once started one cannot stop it. Even if one consciously decides to not perform the process prior to seeing the triggering stimulus, knowing full well that the stimulus is about to be

presented, one still cannot stop its occurrence in the face of this preparation. If I asked you to look at the sky and not perceive its color, you could not do it. The mere presence of the stimulus (sky) triggers the response (color perception) regardless of any goals you might have to prevent the response.

As another example, if I ask you to ignore the meaning of the words written in this sentence and focus only on whether the shapes of the particular letters are curved versus angled, you would still likely extract the meaning of the words. Extracting word meaning occurs even when we consciously try to control it. Stroop (1935) provided a classic experimental illustration of this fact. Research participants were shown either words that named a specific color ("red," "blue," green," etc.) or patches of the color. The words were always printed in colored ink, but the color of the ink did not always match the word. For example, the word "red" might be written in blue ink, while the word "blue" might be yellow colored. The task was extremely simple: Name the ink color. This is facilitated by not reading the words. Naming the ink color is difficult when it is in a word as opposed to a rectangular patch. When the word "red" is written in blue ink, people instead start to name the printed word ("red") rather than the ink in which it is printed (blue). The processing of word meaning occurs immediately upon perceiving the word. They then need to stop and correct themselves, which interferes with the ability to do the task asked of them (naming the color) relative to just seeing a patch of color. Why? Indicating you have seen the color blue when encountering the word "red" in blue ink is in conflict with an uncontrolled response of reading the word (what we usually do with words). The inability to control one response (reading) slows the designated response (naming colors). This **Stroop effect** illustrates interference due to lack of control over an automatic process.

Stroop effect: When automatically detecting a word's semantic meaning is illustrated by interference with another task. For example, if "red" is written in blue ink and the task is to name the ink's color, we can discern that the semantic meaning "red" is automatically triggered if interference with saying "blue" occurs when naming the ink color.

One can use a similar methodology to illustrate that processes in person perception are beyond control. For example, Geller and Shaver (1976) presented words to people and asked them to name the color of the ink. This time, however, instead of the words being color names, the words were either self-relevant to the people who were reading them or neutral words that were not relevant to the research participants. The logic was that stimuli that are relevant to us will be detected and processed without being able to control it. We cannot help reading words that are presented to us, and when those words are relevant to us we find ourselves distracted by them and wanting to linger on them. This increased attention to the word meaning is in conflict with the task. Rather than saying the color of the ink as fast as possible, we are sidetracked by the processing of self-relevant information in our environment. A similar finding was produced by Bargh and Pratto (1986). Words pretested to be part of a participant's self-concept were shown in colored inks, along with words that were not self-relevant. Participants were to name the color of the ink. The reaction times to naming ink colors were reliably slower when the word content (which should be ignored to efficiently do the task) was consistent with the participant's self-concept. Attention was uncontrollably drawn by word content that was self-relevant. As Shiffrin (1988) stated: "If a process produces interference with attentive processes despite the subject's attempts to eliminate the interference, then the process in question is surely automatic" (p. 765).

Lack of Awareness

Lack of awareness of a cognitive process is perhaps the easiest from among the features of automaticity to grasp intuitively. Many of the cognitive tasks we engage in occur without

our awareness; the feeling of gears churning is absent. We can drive while lost in thought, without any awareness of what we did during the last 4 miles. We earlier reported an experiment by Murphy and Zajonc (1993) in which people were influenced by facial expressions in photos that were presented subliminally. If the facial expressions were never consciously seen, the person would obviously not be aware of its influence. Yet an influence was evidenced all the same. In the Murphy and Zajonc research, perceivers are not aware of the processes through which a subliminally presented picture of a face impacts on their evaluation of an object. Indeed, they are not even aware of the existence of the influencing force (the facial expressions).

Varieties of Automaticity

Most information processing contains some subset of these four criteria for automaticity, but not all of them. A running example has been that of driving a car. It does not meet all these features, yet driving does not always occur with full consciousness and one's awareness focused on the task. To call this activity fully under control would seem to misrepresent the process. But so too would calling it automatic. To capture the full complexity of most processes, especially those involved in perceiving other people, Bargh (1989) proposed that there are varieties of automaticity. This allowed for the possibility that the four features of automaticity could appear in various combinations. If a process possessed all four features, such as when one perceives color, it was said to be a particular variety of automatic processing: **preconscious automaticity**.

If one does not intend to initiate a response and even lacks awareness that the response has occurred, but its occurrence requires some type of conscious processing, then a second variety of automaticity is said to exist: **postconscious automaticity**. For example, you may unintentionally and unknowingly find yourself sending nonverbal cues to someone you are interacting with, yet are conscious of the interaction and of the fact that nonverbal signals of some type are being sent. These cues could unintentionally signal to that person that you are uncomfortable around them (e.g., you lean away from them, fail to make eye contact). The perceiver might not realize they are seeking out facial cues and the person whose face it is might not intend to have anything but a neutral facial expression. Yet information about broad personal qualities, such as dominance, power, and trustworthiness, are sent through facial features and are detected by perceivers (Todorov, Said, Engell, & Oosterhof, 2008). Micro-expressions one does not intend to send can reveal when one is lying (Ekman & Rosenberg, 2005). However, such unintended responses would not have occurred had you not consciously decided to interact with the person who is detecting these unintended facial expressions.

Finally, some processes occur without conscious awareness and with great efficiency, but require that one has a conscious goal in place for the response to be initiated. This variety of automatic processing was labeled **goal-dependent automaticity**. A perfect example has already been discussed at length: driving. Driving is efficient (you can go miles without

Preconscious automaticity: A cognitive process that has all four features of an automatic process—it is not consciously intended, it is efficient, it happens outside conscious awareness, and it is unable to be controlled.

Postconscious automaticity: A cognitive process that is efficient and occurs without awareness but that can be controlled when conscious processing demands it. For example, you may unintentionally and unknowingly send nonverbal cues during an interaction, but can control it when desiring to control it. Consciousness allows for control over the process.

Goal-dependent automaticity: A cognitive process that requires one to have a conscious goal for the response to be initiated, but runs outside awareness and without conscious monitoring. For example, you may intend to mentor another person but unintentionally trigger microaggressions when doing so that would have been absent without a mentoring goal.

it being disrupted by simultaneous tasks such as making a call, being lost in thought, and singing quite loudly). But it requires having the goal to drive, and starting and stopping can be controlled by willing it.

EFFICIENCY AND IRRATIONALITY

Identifying the various features of automaticity and incorporating them into the definition is important if one hopes to distinguish an automatic process from other responses people make that are similarly efficient. Why is it essential to make this distinction? Because equating the term "automaticity" with low effort can give rise to the false conclusion that unconscious processing is irrational processing. Automatic processing is often "rational" and accurate, while conscious processing is at times low in effort and irrational. It is important to distinguish automatic processing from processes that simply reflect irrationality and mindlessness.

Mindlessness

Automatic processing is not simply a shortcut people use to avoid thinking deeply. Rather than laziness, it is a routine set of responses that are associated with a stimulus that allow one to develop increased efficiency at a task. It allows one to trigger associations (which may or may not be accurate) once detecting a cue that is diagnostic of the category. For example, one may detect a person wearing a turban, and this might trigger one's associated knowledge of the various cultures in which people wear headwear with cloth winding in this fashion. If one's schema matches that specified by Wikipedia, then one would have knowledge that a variety of cultures have people who wear turbans and this includes communities located in "the Indian subcontinent, Southeast Asia, the Arabian Peninsula, the Middle East, the Balkans, the Caucasus, Central Asia, North Africa, West Africa, East Africa, and among some Turkic people in Russia, as well as Ashkenazi Jews." It need not be an automatic process that causes a person with such a schema to incorrectly categorize a person wearing a turban as Middle Eastern. The automatic process associates turbans with this group, but with many other groups as well. If a person with such a schema were to misidentify another person wearing a turban as Middle Eastern solely on the basis of them wearing a turban, this would be an incorrect use of conscious processing, not a flawed automatic process.

In this example there is an automatic process that triggered many possible groups that are all associated with the category. Yet the error occurs not in associating these groups with the category but in the irrational narrowing of focus on one of these groups to the exclusion of the others when there is no other reasonable evidence to do so, and perhaps good evidence to suggest a different group is more relevant (such as Balkan). The easy triggering of information should not be confused with the low effort use of conscious reasoning to limit what we think. Bargh (1984) makes this point by contrasting automatic processing with **mindlessness**: a type of thinking about other people in which mental energy and effort seems to have been eliminated in how we consciously attend to their behavior and features. Langer, Blank, and Chanowitz (1978) introduced the term, and defined it as responding initiated in a situation when "attention is not

Mindlessness: A type of thinking where mental effort is apparently eliminated and we instead operate using existing knowledge about situations/people. For example, scripts, frames, and schemas specify appropriate ways to act that can merely be triggered by cues in the environment, allowing one to respond without the need for mental elaboration.

paid precisely to those substantive elements that are relevant for the successful resolution of the situation . . . new information is actually not being processed . . . what is meant by mindlessness here is this specific ignorance of relevant substance" (p. 636). Like an automatic process we operate on the basis of existing knowledge about situations and people, but unlike an automatic process that knowledge is triggered because of a faulty and incomplete conscious assessment of the features of the stimulus. Whereas automatic processing reflects unconscious processing, mindlessness involves an incorrect use of conscious processes in an effort to not think deeply. The difference is subtle. Automaticity is when a cue that has, through habit and conditioning, been associated with a representation triggers that representation without one knowing (and perhaps without even knowing the cue is present). Mindlessness is when the same responses are triggered by the same cue, but because one consciously chooses to focus on that cue to the exclusion of other information in the situation that would render the responses inappropriate. A selective ignoring of some information, the choice not to process relevant stimuli in the situation in favor of the simplicity offered by the script, is its hallmark. While automatic processing may make it easier to detect some stimuli over others, this is not the same as choosing to ignore some stimuli because it would be effortful or undesirable to do so. One is about the efficiency of thought, and one is about a conscious choice to not exert necessary effort.

Langer et al. (1978) assumed that mindlessness resulted in behavior produced without the benefit of conscious consideration of the cues in the situation; behavior initiated because a superficial assessment of a situation triggers a script or schema. It is mindless not because the response is automatically associated with a schema but because the schema is triggered by a superficial assessment. Situations we enter into, such as getting a beverage at the coffee shop, have features and cues that tell us how to act. Conscious processing is focused on verifying the script as appropriate for this situation. The script may tell you that first you wait on line, then you order, then you pay, then you walk to the other end of the counter, then you retrieve your beverage. But if you wait behind a group of people only to learn they have already ordered, this is not a problem with automatic processing being irrational. You have simply failed to dedicate enough conscious processing to detect that there is no line, so the script is not relevant in this situation. According to Bargh (1984), with mindlessness "the result is that certain pieces of information are selected by the script over others that may actually be more relevant and useful in the current situation" (p. 35). People scan the environment in a way that fails to detect germane information and avoid using information they should use (if operating rationally).

A well-known experiment by Langer et al. (1978) makes this point using the category of "a favor." If a favor is requested, then the triggering of an associated response would be a type (or variety) of automaticity. There are scripts that specify how to act if a favor is requested. However, if another person does not make a legitimate favor request, it is not an automatic process if you respond with the scripted response. This is simply you, the perceiver, getting the category wrong because not enough attention was paid to realize that the script for "favor" is not appropriate in this instance. This is superficial processing of the features and assigning the wrong label, but then using the right associations to that wrong label. The association would not be irrational had the schema for "favor" actually been invoked. But one's low effort at attending to the situation caused the wrong category to be invoked, and hence the response is irrational. Langer et al. had an experimenter approach unsuspecting people who were using the copy machine at the library. These people were asked if they would step aside and allow the experimenter to use the copy machine immediately. They reasoned that this request could trigger a script in the mind of the person using the copy machine that a favor was being requested, and that favors reflect either an

urgent need, or an emergency. The triggering of the script for "a favor," they reasoned, would depend on whether or not *a reason* was provided by the person making the request. If the person merely said, "please stop what you are doing and allow me to cut in and make copies immediately," without providing any reason, they would likely be seen as rude and the request denied. However, if the person offered a reason, then the script for a favor would be triggered. Langer et al. argued that because of mindlessness, all that was needed to trigger the script was for the person making the request to *seemingly* offer a reason. It should not matter if the reason offered was legitimate or ridiculous. If the person being asked is mindlessly processing the request, they would relinquish the copy machine.

Langer et al. (1978) argued that if any reason offered—even a ridiculous one—resulted in the person acquiescing, then we would have evidence of mindlessness. The person at the copy machine would simply follow the rule of "if a favor is requested and is accompanied by a reason and I am not being burdened in any way, then comply." They would stop assessing whether the reason was legitimate. To show this, some of the people using the copy machine were approached and asked, "excuse me, I have five copies to make, can I use the machine instead of you *because I am in a rush*." Other people using the copy machine were approached and asked, "excuse me, I have five copies to make, can I use the machine instead of you *because I have to make copies*." Each of these requests seem to follow the script for a favor by the person offering a reason for the request. Yet one of these requests offered no reason at all, but just followed the format of providing a reason—can I make copies because I have to make copies. Evidence for mindlessness was found because each "reason" was equally effective at having the request granted; both requests led the person to relinquish their use of the copy machine more than when the person was approached at the copy machine and asked without a reason being provided—"excuse me, I have five copies to make, can I use the machine instead of you." Adding "because I have to make copies" offers no additional information, and is thus not a real reason. But it is effective. Why? Because it triggers mindlessness where people stop paying attention to relevant details and then surrender their action to the scripted response. Even though the script for a favor does not really apply. They irrationally treat a person with a nonsensical and vapid "reason" better than a person with no reason. It is not the script and the process of triggering responses associated with it that is irrational, but the mindlessness on the part of the perceiver consciously invoking the wrong category and hence the wrong script. Mindlessness is a poor deployment of conscious processing, not irrationality of the automatic processes. It exists when people act and think with a low level of conscious involvement and end up not making use of (or paying attention to) all the relevant details in their environment. To quote Langer et al. (1978): "only a minimal amount of structural information may be attended to and that this information may not be the most useful part of the information available" (p. 641).

Are Conscious Processes More Irrational Than Automatic Processes?

Research on mindlessness points out that consciousness does not guarantee that people will think rationally, and research on automaticity highlights that not all processes that lack conscious awareness are irrational. Automatic processing is not to be equated with irrational outcomes. When an automatic process leads us to retreat into the unconscious it can service our ability to detect relevant information in the environment, just as conscious processing can lead us to ignore relevant information. Huang and Bargh (2014) argued that in many areas, such as self-regulation, automatic processes were an evolutionary earlier development relative to conscious processes; during self-regulation people often operate better without the burden of consciousness. Conscious processing can yield undesired

results, and an implicit form of regulation might avoid these errors and biases. Typically, a process becomes automated through practice, and that practice is engaged in because the process is making responding easier and more efficient. Irrationality is not a feature of automaticity, but efficiency is.

For example, a conscious goal to control the use of stereotypes can have unintended consequences that promote stereotyping (e.g., Macrae, Bodenhausen, Milne, & Jetten, 1994). Asking people to explicitly try and be colorblind leads to them using race more rather than less (Norton, Vandello, Biga, & Darley, 2008). These acts of consciousness result in the opposite outcome. As another example, telling people to be creative, and giving them direct instructions about what not to do (e.g., do not plagiarize, do not copy the names of existing products when generating new brand names), leads to a lack of creativity. People in such experiments plagiarize more and copy brand names to a greater degree (e.g., Marsh, Bink, & Hicks, 1999; Marsh, Ward, & Landau, 1999; Smith, Ward, & Schumacher, 1993). Wegner and Erskine (2003) describe an entire class of unintended effects that result from people trying to exert control over their own unwanted acts. Conscious thought is not always better than automatic processing.

There are many lines of work that illustrate that at times the elimination of consciousness creates better efficiency, a reduction in bias, and better outcomes (e.g., Dijksterhuis, Bos, Nordgren, and van Baaren, 2006; Dijksterhuis & Nordgren, 2006; Sassenberg & Moskowitz, 2005). Let us return again to the example of being creative, which by definition requires that one avoid conventional ways of thinking and typical associations. Research has shown, as noted above, that asking people to try and be creative has the opposite effect. However, when creativity is triggered outside of conscious awareness and the pursuit is turned implicit, the desired outcomes of heightened creativity are achieved (Sassenberg et al., 2021; Sassenberg & Moskowitz, 2005). Similar benefits of unconscious cognition can be seen in decision making. Dijksterhuis et al. (2006) gave some participants a conscious task of choosing between several products (e.g., from among four apartments to rent). They were given time to consciously evaluate the qualities of each option and then make a choice. Other participants were not given time to consciously evaluate the qualities that differentiated among the options. They found that the people who were not able to consciously deliberate made objectively "better" choices than people with conscious effort exerted. The logic of this research is that when denied the chance to consciously evaluate, people still had the goal of making a choice and continued to deliberate and assess the options outside consciousness. The processes used to regulate this goal were happening outside awareness and were less prone to bias than the conscious ones.

In another illustration of the benefits of automatic processing, Shah (2003) showed that when people had a goal of performing an analytical reasoning task that required conscious effort, performance on that task was facilitated if a second goal had been unconsciously triggered in the same participants. This was only true if the two goals were compatible with each other (such as an unconscious goal to be creative). Thus, whereas trying to meet two conscious goals would be effortful and overloading, trying to meet an implicit goal is not an overload to the conscious goal. It can actually make a conscious goal easier to reach. People are more efficient when a single behavior can serve multiple goals that can each help toward performance (e.g., Chun, Kruglanski, Friedman, & Sleeth-Keppler; 2011). Fishbach, Friedman, and Kruglanski (2003) showed that when an unconscious goal was incompatible with a conscious goal it could still produce more efficient responding by inhibiting the conscious goal (especially if the unconscious goal was the more important). For example, people with an unconscious goal of eating healthy were able to inhibit their conscious goals relating to pleasure eating that were harmful to them in the long term.

Macrae, Milne, and Bodenhausen (1994) argued that if automatic processes are to be thought of as promoting efficiency and better responding, there should be demonstrable benefits; information processing should be easier, more efficient, and cognitive resources preserved when automatic processing accompanies a conscious task. For example, even though we think of stereotypes as negative, if they evolved to make us more efficient, then they could serve as a useful means for economizing cognition. People who use stereotypes on a task should be able to think less and arrive at decisions about people more quickly than people who do not (e.g., Gilbert & Hixon, 1991). This "savings" afforded by the use of the stereotype should be reflected elsewhere with increased efficiency; they should be better at another task they perform at the same time. Macrae et al. found that people who unconsciously used a stereotype in an impression-formation task had attentional resources liberated that were then used to assist in executing a reading comprehension task that required intense focus. People without a stereotype to use on task 1 performed worse on task 2.

Throughout this book we see many examples of automatic processing producing unwanted outcomes and biases. Indeed, this is how many people think of unconscious processes: either through the Freudian lens of trauma or this more modern lens of bias. When bias is produced by automatic processing, conscious processing can help to overcome those errors and mitigate the bias. Our point here is that this view is unbalanced. Both automatic processing and conscious processing can produce better cognitive outcomes, and both are capable of leading to bias. The argument put forth here is that automatic processing is the child of desires for efficiency and ease, and not of trauma and error. Even some of the errors produced by automatic processing are beneficial to the organism. Balcetis and Dunning (2010) showed that people perceive things that they are motivated to acquire as closer to them than objects they do not desire. Thirsty people perceived a bottle of water as 1.1 times closer. Less wealthy college students saw a $100 bill they could win as 1.2 times closer than a $100 bill they could not win. People who felt strong disgust for insects perceived a spider to be 1.5 times farther away than people with no such aversion (Cole et al., 2013). Women who saw a man urinating in public perceived him to be 1.4 times farther away than an angry man (Cole et al., 2013). When people need to act to acquire a desired reward or avoid danger, their automatic processes bias their perception of distance. But this bias is helpful to them and serves their well-being.

AUTOMATIC ATTITUDES AND BELIEFS: THE IMPLICIT NATURE OF IMPRESSIONS OF GROUPS

> The capacity to evaluate other people is essential for navigating the social world. Humans must be able to assess the actions and intentions of the people around them, and make accurate decisions about who is friend and who is foe, who is an appropriate social partner and who is not. Indeed, all social animals benefit from the capacity to identify individual conspecifics that may help them, and to distinguish these individuals from others that may harm them. Human adults evaluate people rapidly and automatically on the basis of both behavior and physical features.
>
> —Hamlin, Wynn, and Bloom (2007, p. 557)

What Is an Attitude?

An **attitude** is an evaluation of a person or an object encountered in the social world—an assessment of the target as positive or negative. Attitudes specify how we feel, dictate our positive and negative reactions, and this naturally guides our behavior (though the links between attitude and behavior are a complex topic that is best reviewed by a more

specialized book or chapter; e.g., Albarracin, Johnson, & Zanna, 2005; Fazio, 1986, 1990b). As such, evaluations orient the person for interacting with the social world, so that the person approaches or engages the stimuli they view favorably and avoids or disengages with the stimuli they view negatively (e.g., Eaton, Majka, & Visser, 2008). Due to their functionality, attitude activation is seen in the psychological literature as among the most primary and important mental activities in which humans engage. Even infants show a preference for a character who helps others and an avoidance of a character who hinders others. If shown animated characters who help another character up a hill, versus characters who prevent another from trying to get up that hill, a 10-month-old will like those characters who help and dislike those who hinder. When presented with both characters, infants choose to play with the one they had seen help (e.g., Hamlin et al., 2007). From infancy we are *evaluating others based on their actions and intentions*. It is not just people but any object, place, or thing is also evaluated. The target of these evaluations is known as an **attitude object**. Fazio, Sanbonmatsu, Powell, and Kardes (1986) argued that people spontaneously retrieve stored attitudes from memory when an attitude object is seen, and that they use this initial attitude as the basis to evaluate, judge, and make decisions about the current attitude object—that is, the attitude is associated with the attitude object, so that when the object is encountered the attitude now has heightened **attitude accessibility**. Once accessible, the association of attitude and attitude object is reinforced.

Attitude: An evaluation of an object as positive or negative. Many theories specify it is more than "affect," it is a structure in memory that also specifies how to appropriately respond to an attitude object (behavior), as well as related knowledge and beliefs (cognition) about the person/thing being evaluated.

Attitude object: A stimulus (person, place, or thing) that is the target of a person's evaluation; the thing to which they have an attitude.

Attitude accessibility: The attitude object serves as a prime that makes the existing attitude have heightened accessibility. Once accessible, an attitude can reinforce its association to the attitude object by being applied yet again to the evaluation of the attitude object that triggered it.

Cognitive response: When attitudes are formed or changed by actively engaging existing beliefs and an internal dialogue regarding what one knows about the attitude object is triggered. This can include thoughts that affirm the existing attitude, but also counterarguments that create doubt about the existing evaluation.

Attitude strength: The degree to which an attitude is important to a person and richly embedded in a network that ties it to one's knowledge, committed goals, social identity, moral beliefs, and vested interests.

Many theories of attitudes specify that an attitude is more than just the "affect" experienced toward a stimulus. The attitude is thought of as a structure represented in memory that contains this evaluative or affective response, but also specifies behavior relevant to how one might appropriately respond to the attitude object, as well as related cognition—knowledge and beliefs—about the attitude object. For example, Greenwald (1968) showed that when an evaluation of an attitude object occurs there is an accompanying **cognitive response**. People actively engage their existing beliefs when evaluating, and the content of this internal dialogue regarding what they know about the attitude object is their cognitive response. This can include thoughts that affirm the existing attitude, but also counterarguments that create doubt about the existing evaluation. And as reviewed in Chapter 2, attitudes can contain associative learning that connects it to affect, but also to propositional learning (e.g., De Houwer, 2014a, 2014b; Gawronski & Bodenhausen, 2018; Kurdi & Banaji, 2023) that dictates the relationship among attitude objects.

Attitudes differ from one another not only in the affective response they trigger but in how strongly those reactions are held (e.g., Fazio, 2007). **Attitude strength** is determined by factors such as (1) how important the attitude object is to the person (its association to goals to which one is highly committed and to the vested interests of close friends, family, and social in-groups), (2) its embeddedness among one's value and belief system (its association in an interconnected network of one's philosophical, political, moral, and religious

values), (3) how well-informed the person is when developing the attitude (more knowledge when forming an initial attitude will lead to it being more strongly held), and (4) its association to social identity concerns. Fazio et al. (1986) illustrate that the more strongly held the attitude is, the faster and easier it is activated from memory when the attitude object is encountered, making stronger attitudes more likely to be reinforced through repeated use. Thus, stronger attitudes can be detected by their ease of accessibility. For example, attitude objects associated with morality should be linked to strongly held attitudes and trigger affect and inferences about whether a person is honest and trustworthy (e.g., Brambilla & Leach, 2014). This is supported by functional magnetic resonance imaging (fMRI) research that shows that the association of trustworthiness with a face (as an attitude object) shows corresponding activity in the amygdala (Winston, Strange, O'Doherty, & Dolan, 2002), which is active when detecting dangerous and threatening stimuli (e.g., Todorov, Dotsch, Porter, Oosterhof, & Falvello, 2013; Todorov, Dotsch, Wigboldus, & Said, 2011).

Are Attitudes Automatic?

Bargh, Chaiken, Govender, and Pratto (1992) argued that attitudes are more than easily made accessible, but that there are **automatic attitudes** (e.g., Bargh, 2017; Fazio, 2007)—activation occurs whenever the attitude object is encountered without the conscious intention to retrieve the attitude. The process of evaluating the stimuli (especially people) we encounter, due to its frequency and repetition, comes to be *automated*. Bargh et al. proposed that people have attitudes relating to *everything* they encounter, and those attitudes are triggered within *milliseconds* of having encountered whatever the thing may be. As described in the quote from Hamlin and colleagues (2007) that starts this section, this triggering of affect is described by researchers as having a functional and adaptive value, and it is this functionality that is believed to be the cause of its habitual use and ultimate automation. And this automatic attitude activation is not dependent on attitude strength. Even weakly held attitudes are activated automatically. Attitude strength might impact whether an attitude guides judgment and behavior, but accessibility of the attitude—its associated evaluation being triggered and made ready to use—depends only on the presence of an attitude object.

Automatic attitudes: Positive or negative affective evaluations that are immediately triggered by stimuli; the mere presence of the stimulus leads to activation of an evaluative response associated with the stimulus. Such implicit affective responses allow us to know if the stimulus is a threat or an opportunity; whether to approach or avoid.

Chen and Bargh (1999) provide an illustration that nicely shows the automatic triggering of attitudes, and the adaptive nature of attitudes via their link to approach and avoidance goals. They argued that if an attitude is activated automatically, then people should be faster to respond in a manner consistent with the attitude, such as approaching a positive attitude object and avoiding a negative one. They reasoned that approach motivations and liking are associated with pulling something toward you, such as with a response that involves an arm flexion. Avoidance and dislike are associated with pushing something away, such as with a response that involves an arm extension. Therefore, if a positive attitude is implicitly triggered, it should make one faster to flex. If a negative attitude is triggered, one should be faster to extend. They asked research participants to simply move a lever when a word appeared. Half of them were told to move the lever toward them (a flex) and the other half were told to push the lever away from them (an extend). They then manipulated whether the words that appeared were positive or negative. Although this task was not ostensibly about attitudes, responses were faster to the positive words (vs. negative words) when participants had to move the lever toward them when the word appeared. The

opposite was true when the arm movement was extension, people responded faster to negative words. The words automatically triggered an attitude, and this activation is reflected in the way approach and avoidance behaviors were facilitated.

Automatic attitude activation even occurs if we do not consciously know we have even seen anything. As Bob Zajonc (1980b) said, "preferences need no inferences" (p. 151). When images of, say, the Pope are flashed at people so fast they cannot report having even seen an image, let alone the Pope, the positive (or negative) attitudes associated with the Pope will be triggered in one's mind (e.g., Baldwin et al., 1990). We evaluate everything, and we evaluate everything immediately, even without knowing we are doing it, and even without knowing we have even seen anything to evaluate. However, the automatic triggering of an attitude should not be confused with a lack of awareness that an attitude exists—that is, one can have explicit awareness of the attitude (e.g., that one dislikes cauliflower) without awareness of how it was formed or how it is activated, or that it was activated in any given moment. Fazio and Olson (2003) point out that a perceiver may not have separate structures for implicit attitudes and explicit attitudes. One might have awareness of the experience of the evaluation itself, but could still lack awareness of the associations that exist in that structure that produce that conscious experience. This means one can have an attitude triggered without realizing how or why, but still consciously experience the negative or positive affect (e.g., Hahn, Judd, Hirsh, & Blair, 2014; Phillips & Olson, 2014). It is also possible for one to *not* experience the affective response, so that the entire process of evaluating is automatic.

Where do automatic attitudes with such easily triggered associations come from? As with learning more generally, attitudes can be learned through conditioning. An object that is novel or evaluatively neutral can develop a positive or negative evaluative association if it is repeatedly paired with an attitude object that has valence (e.g., Crano & Prislin, 2006; Eagly & Chaiken, 1993; Hofmann, De Houwer, Perugini, Baeyens, & Crombez, 2010; Olson & Fazio, 2001, 2003, 2006). The neutral stimulus is known as the conditioned stimulus (CS), and the positive or negative object with which it is repeatedly paired is known as the unconditioned stimulus (US). The CS comes to take on the valence of the US and an attitude has been formed through a process known as **evaluative conditioning**. This process may also involve the encoding of propositional information that describes relationships among the stimuli (e.g., De Houwer, 2018; De Houwer & Hughes, 2016).

Evaluative conditioning: A stimulus-driven process in which an attitude is formed through repeated pairing of an object with an unconditioned stimulus (US) that has either positive or negative affect. Through this pairing the conditioned stimulus forms an association to the US in long-term memory, and is linked to its affect.

Is Prejudice toward a Group of People Automatic?

The automaticity of attitudes extends to all people and objects for which we have mental representations. This means we have attitudes toward groups of people as well. **Prejudice** is such an attitude. It is a prejudgment of a group of people, usually a negative one, where *evaluations* of a group are held in our mental representation for that group, and typically applied to individual members of the group. Just like any attitude, the attitudes toward a group can develop through conditioning, and these associations can be triggered automatically. For example, prejudice toward a group of people could be conditioned by a person forming associations in memory among the group (that might start out as neutral

Prejudice: A shared prejudgment of a group, usually negative. It is an association between *evaluations* of a group—positive or negative attitudes—with a mental representation for that group. If triggered, these attitudes are typically applied to individual group members, with little awareness of the attitude accessibility or its influence.

or unknown) and negative behaviors or traits (such as violent, criminal, unintelligent, or generally bad) with which the group is associated via repeated exposure. This can happen unintentionally, or intentionally, through selective media exposure. Weisbuch, Pauker, and Ambady (2009) found that when participants were exposed to television programs portraying White characters expressing negative nonverbal behavior to Black characters, the participants subsequently had increased prejudiced to the group "Black people." Similarly, Lamer and Weisbuch (2019) found that participants who saw images of men consistently placed in a higher position on the page than images of women associated men with dominance. Prejudiced attitudes were conditioned.

Once such associations among negative evaluation/affect and a specific social group develops, it remains possible that such associations become triggered outside conscious awareness. A wide variety of measures now exist that allow us to illustrate the "automatic" activation of prejudice. Though people would deny it explicitly, these measures reveal that research participants associate positive reactions with groups to which they belong and negative reactions to groups that are stereotyped in their culture. For example, Fazio, Jackson, Dunton, and Williams (1995) presented an attitude object followed by a positive or negative adjective. The participant was to respond whether the adjective was "good" or "bad," with the speed of this response facilitated if affect had previously been triggered by the attitude object. In this case the attitude object was a photograph of either a White or a Black man's face. The adjectives were positive (e.g., attractive, likable, wonderful) and negative (e.g., annoying, disgusting, offensive) words that are irrelevant to the stereotypes of these groups. This allows a test of whether positive or negative affect is immediately triggered upon seeing the face. The findings revealed that White participants were faster at making the evaluations when positive adjectives were preceded by faces of White men and negative adjectives were preceded by faces of Black men.

Wittenbrink, Judd, and Park (1997) used a similar procedure, except White participants saw the *subliminal* presentation of a group label ("Black" vs. "White") rather than a face. Also, instead of judging whether an adjective was "good" or "bad," participants simply were asked to indicate whether a string of letters was a word (a lexical decision task). When the string was in fact a word, it could be either positive or negative. The results revealed that participants responded faster to positive words when they followed the label "White," yet they responded faster to negative words when they followed the label "Black"—despite the labels being presented outside of conscious awareness. And the effect was strongest when the words were stereotypical of the group. An affective response was triggered by the mere presence of the group label, with participants remaining unaware of either the label's presence or their affective reaction to it.

In the next chapter, we focus on a variety of measures that have been developed to assess implicit cognition, including many that have been used to reveal prejudiced attitudes. We end this discussion of prejudice by noting that we have not really answered the question posed by the heading to this section. Is prejudice automatic? What we have seen is that people often do not realize they have prejudice, and if they are not aware of it, this makes control over it less likely. You do not have to be an explicit racist to be biased and make racist assumptions and choices. That is what preconscious influences do to you—it seems like the person *is* this way but it is really just an assumption; it is information added in by attitude activation. As a perceiver we feel as if the information came in from outside, via our senses, because it is fast and efficient. But does this efficiency and invisibility make our prejudice automatic? The answer to this is complicated, since we have defined a variety of types of automaticity. To fit the definition of preconscious automaticity these processes would need to be uncontrollable. In Chapters 11 and 12 we review evidence that argues

these affective reactions—though outside awareness and efficient, and perhaps triggered silently by cues—are controllable. While more difficult to control than an explicit reaction, even a process that is fast, efficient, and implicit can be controlled. In this way prejudice is not fully automatic but goal dependent.

CONCLUSIONS

In psychological science, there is an emphasis on individual reporting as a primary method of discovery about the nature of cognition. There are modern tools such as fMRI and event-related potential (ERP) that allow insight about what is happening in the brain, but since we are studying human cognition, our most common course of action is to ask those humans to report about their own cognitive experience. This is not problematic for some areas of investigation. If our interest is in the perception of two very similar stimuli, we can trust the individual to report whether they notice a difference between them. For example, which stimulus is faster or is brighter? If we are interested in the selective nature of attention, we could flash an array of stimuli at a perceiver and ask them to report what they can recall having seen from among the set. In each of these examples, the report might be inaccurate, but there is little reason to suspect the person is not reporting what they believe to be true of the stimulus. Self-report can be a trusted tool if our concern is with what people consciously see and recall (and they are motivated to be honest). However, that trust starts to erode if our concern shifts from what people see and recall to how people know what they see and recall. Can people accurately report on *how* they think? Can people accurately report on the mechanisms of cognition?

As you know from personal experience, people lie to others during self-report for a variety of reasons that allow them to look good and be accepted: ingratiation, conformity, manipulation, flattery, kindness, rhetoric, and so on. When people self-report they are often not reporting what they actually think but creating a socially desirable impression in that moment. The most generous interpretation is that they do not know what they truly think and are simply reporting what is most salient to them at that moment. Their processing is automatic and they cannot see it, so they report as best as they can. But it is inaccurate. A less generous interpretation is that people lie to others so they can look good. Being seen as biased is especially threatening to people who believe they are not, and the fear of being mislabeled as a biased person causes anxiety and arousal in such people.

Of course, people also lie to themselves, seeing themselves as more unbiased than perhaps reality dictates. This can be seen in the disjunction between self-reports about their own bias and more subtle measures, as well as in the different types of behaviors predicted by each of these measures of bias. For example, direct and overt ratings of prejudice can reveal low amounts of bias when in the same person more subtle measures of bias (which we review in the next chapter) can reveal high levels of bias. The explicit and implicit attitudes diverge. Additionally, the overt measure is correlated with explicit and deliberative behavior, such as what a White person says when interacting with a Black person (such as promoting the legitimacy of anger in the Black community). But the indirect and implicit measures of the same attitude are correlated with spontaneous acts and nonverbal behavior that signals avoidance and aversion (e.g., Dovidio et al., 2002; Fazio et al., 1995). Because people are less able to monitor their nonverbal (and other forms of spontaneous) behavior, such behaviors are often seen as more honest, and relied upon more heavily than what people say and do with more deliberation. When the two do not align, the deliberate behavior is seen as untrustworthy.

An automatic process is not only one that people have difficulty seeing and monitoring but is also one that they cannot control, and never intended to initiate. For example, categorizing another person's facial expressions as happy or sad will happen upon mere exposure to that expression. One does not need to want to infer their state, nor can one stop oneself from knowing what emotion is being communicated. Similarly, having an attitude does not require a request to form an opinion. The mere detection of an attitude object triggers the corresponding affective response. When people process automatically this is not the same as saying they are lazy, or irrational, or thoughtless, or thinking poorly, or lacking a desire to get it right. Automatic processes can be efficient, and can avoid errors that conscious thinking would produce (e.g., Hasher & Zacks, 1979; Shiffrin & Schneider, 1977).

Importantly, automatic processing and control are not opposites. Lack of control is an element of automatic processing, but not the only defining feature of automaticity. So, a process over which one has control is not correctly called the opposite of automatic thought. People too readily see these two constructs as endpoints of a continuum of processing: automatic and controlled. They actually address different things. Control relates to one's ability to self-regulate. People typically use it to mean to exert conscious effort to regulate. However, as discussed above, self-regulation is at times better when it lacks consciousness. The idea of unconscious control is counterintuitive to most people, but you should be able to call upon examples of it from your own life. Many people report that when driving home they need to go to the restroom as soon as they hit the street where they live. This is no coincidence. They had been controlling this urge unconsciously, and cues that signal "home" signal an opportunity to act on it. Examples of such invisible control are around us all the time. We will suddenly see the mailboxes on a street we walk every day when we have the rare need to mail a physical letter. The mailboxes were not invisible. You just have current goals that lead you to seek them out even though you are not consciously seeking them out. Just as with the cocktail party effect, our goals control how we scan the environment and what reaches consciousness. To control is to self-regulate or for a process to be modifiable by intentions. But those intentions and goals do not need to be consciously initiated, the control exerted does not need to be aware to you, and control need not be characterized as lacking efficiency or being effortful. It can be either effortful or effortless. It can be exerted with or without awareness. It is not the opposite of an automatic process. It is a feature of an automatic process. This gets confused when people incorrectly define an automatic process as a lack of awareness and effort, and control as the effortful attempt to reach some goal of which one is fully aware.

The unconscious is not a warehouse for trauma (though it can be) but a tool to produce functioning cognitive responding in a complex social world. Abelson (1981) provided an iconic illustration of how common situations are processed mindlessly, with people following a standard script for how to act in that situation without needing to engage conscious thought. Scripts and schemas were said to be helpful in guiding our behavior in positive and useful ways, allowing people to rely on past experience to guide appropriate action in the moment. Miller, Galanter, and Pribram (1960) stressed that if we had to do everything consciously and deliberately we would never be able to get out of bed in the morning, rendered immobile by having to control each and every muscle with our limited processing capacity. This brings us back to William James's axiom that consciousness drops out of any process where it is no longer needed. Adults forget how difficult many of our hard-earned skills are—we take them for granted. As John Bargh recently told me,

> "I just taught my daughter how to drive a car and was reminded again that what seems so easy for us is really difficult when you are learning—there is enormous savings and

reduction of strain on conscious resources from experience and practice and so much of the activity is done 'for us' by these automatic processes."

Social cognition has allowed us to figure out how to study such important but invisible phenomena that are so central to social life. We delve into some of the methodological accomplishments for studying invisible processes in the next chapter.

NOTE

1. Wundt himself did not rely solely on introspection and was instrumental in ushering in an age of methods used to explore unconscious thought, methods we introduce in this chapter's section on "Automatic Attitudes and Beliefs" and review in detail in Chapter 4. For example, Feldman Barrett (2009) stated that Wundt "invented the reaction time experiment to measure the speed of perception by presenting participants with a tone or light of a particular color and measuring their latency to press or release a button in response. With these first experiments in psychology, Wundt's goal was to identify and measure the atoms of the mind—the most elemental processes that are the basic ingredients of mental life" (p. 314).

Guilford Publications
370 Seventh Avenue
New York, NY 10001
212-431-9800
800-365-7006
www.guilford.com