Sociological theories have long sought to understand the role cognition plays in social process (e.g., Bourdieu 1990; Giddens 1984; Lizardo 2017; Vaisey 2009). In recent years, a “dual-process” model of cognition has emerged as a candidate framework for studying the sociological implications of cognitive processes in a rigorous fashion. Supported by a large body of empirical work in psychology and other cognitive sciences, the dual-process framework posits that there are two general types of cognition (Lizardo et al. 2016). The first type is characterized by automatic execution, and is often fast, effortless, and occurs with little conscious awareness (hereafter automatic cognition). The second type is characterized as intentional, more effortful, conscious, and executes more slowly (hereafter deliberate cognition). Dual-process approaches have yielded new insights into important sociological questions, such as the role of motivation in explaining behavior (Chaves
Despite their rich potential, contributions from dual-process models in sociology have been limited mainly to theoretical innovations, with empirical research lagging behind. The primary reason for this, we believe, is methodological. Although sociologists have ready tools for accessing deliberate cognitive processes—such as forced-choice surveys and interviews—agreed-upon measures of automatic cognition are lacking. The absence of established and valid measures of automatic cognitive processes makes it difficult for empirical research to keep pace with theoretical innovations and evaluate theoretical claims. As a result, many theoretical assertions—for example, about the impact of automatic categorization in reproducing inequality or the role of childhood socialization in shaping educational attainment (Lamont and Molnár 2002; Lareau 2003)—are evaluated indirectly.

In this article, we test the validity of several proposed measures of automatic cognition. We begin by briefly outlining how sociologists have made use of automatic cognition to shed light on various social phenomena, focusing on social stratification and social action. We then discuss the problems inherent in current sociological measurement practices and argue that implicit measures offer a rigorous alternative. In particular, we consider measurement strategies based on the Implicit Association Test (IAT) and the Affect Misattribution Procedure (AMP). Although both of these strategies were originally designed to capture implicit attitudes, they can be adapted to measure a range of constructs, giving them a wide field of application in sociology (Miles 2019). To determine how well each approach captures automatic processes, we examine how well the AMP and the Brief IAT (BIAT)—a simpler measure derived from the IAT—predict automatic influences on behavior during a survey task. To ensure our results neither capitalize on chance nor are specific to any particular social domain, we repeat our analyses in three samples focused on politics, morality, and racial attitudes. In short, we offer sociologists a clear method for measuring automatic cognitive processes so they can focus on using them to answer important theoretical and substantive questions.

**AUTOMATIC COGNITION IN SOCIOLOGY**

Explanations relying on automatic cognition are common across sociological subfields. Although most sociologists do not explicitly use dual-process models, many sociological arguments implicitly draw on the idea of automatic cognition by proposing mental processes that are implicit, unconscious, instantaneous, habitual, uncontrollable, or emotional (e.g., Quillian 2008; Ridgeway and Krchel-Katz 2013; Rivera 2015). Under such guises, automatic cognition plays a key role in theories seeking to explain fundamental sociological phenomena, such as social stratification along axes of gender, race, and class.

Ridgeway and colleagues, for example, argue that primary cultural frames like gender and race are used “automatically,” “near instantaneously,” and “unconsciously” to interpret information and behavior in interactional settings (Ridgeway and Krchel-Katz 2013:297–300; cf. Ridgeway 2011). The cultural meanings surrounding these primary frames—for example, how men and women should act, which skills come naturally to them—“shape interpersonal hierarchies of influence, status, and perceived leadership potential in ways that reproduce and maintain gender and race as systems of inequality” (Ridgeway and Krchel-Katz 2013:300). Automatic cognitive processes are thus argued to play a key role in reifying social differences by shaping individuals’ perceptions of others.
Similar arguments appear in the race literature under headings such as “unconscious racism,” “implicit prejudice,” or “white habitus” (Bonilla-Silva 2003; Quillian 2008). Negative automatic reactions to non-whites have been suggested as (partial) explanations for why black men are more often reported as suspicious by residents in a community (Lowe, Stroud, and Nguyen 2017), why police seem to view black protesters as especially threatening (Davenport, Soule, and Armstrong 2011), and why defendants with stereotypically black features are given harsher sentences (King and Johnson 2016).

Automatic cognition is also readily apparent in research on education. Drawing on Bourdieu, education scholars argue that class-based experiences in early life have lasting effects on how children and youth interact with educational institutions. Children possessing the right dispositions and interactional skills—which Bourdieu calls *habitus* or embodied cultural capital—navigate education systems more easily, resulting in greater opportunities and higher achievement (Beckman et al. 2018; Gaddis 2013; Lareau 2003; Weininger and Lareau 2018). Crucially, Bourdieu conceives of *habitus* as being deeply ingrained such that it operates pre-reflectively and “without presupposing a conscious aiming at ends” (Bourdieu 1990:53; cf. Bourdieu 1984:466)—that is, via automatic cognition.

Automatic cognition has an especially large presence in the sociology of culture, where scholars argue that it is one of the primary means by which culture influences action. As a result, sociologists of culture have been enthusiastic about adopting dual-process models (Lizardo 2017; Lizardo et al. 2016; Vaisey 2009). Evidence from this research suggests that deeply internalized or non-linguistic forms of culture shape action both by influencing the thinking processes that precede behavior (e.g., perception, interpretation, and motivation) and by facilitating certain types of activity (e.g., habits, motor-schematic competencies)—what Lizardo and colleagues (2016) call “culture in thinking” and “culture in action” processes (Chong and Druckman 2007; D’Andrade 1992; Forster and Liberman 2007; Gantman and Van Bavel 2014; Gross 2009; Hunzaker 2016; Joas 1996; Lizardo 2017; Miles 2014a; Neal et al. 2012; Wood and Rünger 2016).

The precise mechanisms linking internalized culture to social behavior, however, remain disputed. Some sociologists of culture (and cognitive scientists) emphasize the primacy of automatic cognition (Evans and Stanovich 2013; Gross 2009; Haidt 2012; Joas 1996; Srivastava and Banaji 2011; Vaisey 2009), whereas others place greater weight on deliberate cognition by emphasizing the interaction of deliberate and automatic processes in producing behavior (Cerulo 2018; Cunningham et al. 2007; Leschziner and Green 2013; Strack and Deutsch 2014; Vila-Henninger 2015). Resolving this issue has important implications not only for sociological explanations of action, but also for understandings of human agency (Hitlin and Johnson 2015) and practical problems such as how dual-process models can inform law or policy (Davidai, Gilovich, and Ross 2012; Ewoldsen, Rhodes, and Fazio 2015; Hitlin and Johnson 2015; Mischel 2014).

**THE MEASUREMENT PROBLEM**

The foregoing discussion demonstrates that automatic cognition is deeply implicated in central lines of sociological inquiry. Sociological uses of automatic cognition, however, are often restricted to theory-building efforts (Gross 2009; Joas 1996; Lizardo 2017; Lizardo and Strand 2010; Vila-Henninger 2015) or empirical studies that infer automatic cognition from its (assumed) indirect observable implications, such as effects on speech or behavior (e.g., Bonilla-Silva 2003; Hoffmann 2014; Lizardo 2017). Automatic cognitive processes are rarely precisely measured, making it difficult to test competing explanations and refine theories (Moore 2017). The reason for this, we believe, is primarily methodological. Measuring automatic cognition requires physiological or rapid-response capturing.
techniques that are difficult to adapt to the observational, interview, and survey methods commonly used in sociological research (for an exception, see Srivastava and Banaji 2011). As a result, sociologists interested in measuring automatic cognition have relied on a variety of ad hoc methods that have not been rigorously validated.

Ignatow (2009), for instance, used “embodied” metaphors taken from online forums as a measure of automatic cognition, while Vaisey (2009) and Gaddis (2013) used traditional forced-choice survey measures. Martin and Desmond (2010) and Moore (2017) used response times, and Pugh (2013) extracted schematic information from interviews. These approaches to measurement have been useful for showing the relevance of dual-process models for sociology. Yet because these approaches are not designed to capture automatic processes, their ability to provide rigorous and reproducible means to measure the cognitive mechanisms of interest remains unclear. This problem is acknowledged with regard to data from explicit self-reports (Moore 2017; Srivastava and Banaji 2011; Vila-Henninger 2015) but is often overlooked for response time data, which are sometimes seen as a solution to the problem (Moore 2017). Even response times, however, do not provide unambiguous evidence of automatic processing, because rapid response can also be produced by shallow or well-practiced deliberate thinking (e.g., computing simple sums [Evans 2012; Stanovich and Toplak 2012]).

More generally, any effort to capture automatic cognition faces significant methodological issues. Several cognitive scientists argue that the defining feature of automatic cognition is that it operates regardless of conscious intent (i.e., it is “autonomous” [Evans 2012; Miles 2019; Stanovich and Toplak 2012]). This means automatic processes are activated without conscious intent when relevant cues are encountered, but it does not mean automatic processes will necessarily run their course and trigger behavior unimpeded. In most dual-process models, deliberate cognition can interrupt or otherwise interact with automatic processes (Evans 2008). As a result, efforts to capture automatic cognition in any of its downstream forms (like talk or behavior) risk capturing deliberate influences in addition to the initial automatic impulse (Vila-Henninger 2015).

A popular strategy for dealing with this problem is to use measurement methods restricting individuals’ deliberate control over their responses—for example, by requiring respondents to engage in a cognitively difficult task while responding (Bargh and Chartrand 2000; Miles 2015). Methods commonly used by sociologists, however, do not usually constrain deliberate cognition. The degree to which much sociological research actually captures automatic cognition is therefore unknown and can be difficult to determine after the fact (cf. Moore 2017).

For example, some studies attempt to infer automatic cognition from observations and interviews (Jerolmack and Khan 2014; Pugh 2013). These types of data necessarily yield downstream speech and behavioral data that likely reflect both deliberate and automatic processes (Lofland et al. 2006; Spradley 1980; Vila-Henninger 2015). Situations that restrict cognitive control might naturally arise during the course of observation (e.g., respondents might be distracted or under time-pressure), but contriving these situations is generally outside the control of researchers. Detecting such situations when they naturally occur is likewise challenging because the processes of intent and control are not observable and must be deduced from (sometimes ambiguous) behavioral data.

Interviews and forced-choice surveys likewise place no constraints on the amount of intentional control individuals exert when providing responses. Respondents might answer questions rapidly, using little deliberate effort and relying on their “gut feelings,” all of which are consistent with automatic processing (Vaisey 2009). However, nothing in these approaches requires this type of response. Individuals completing forced-choice items might also reason carefully, consult their memories, or edit their answers,
processes that are governed by deliberate cognition (Moore 2017; Srivastava and Banaji 2011; Vila-Henninger 2015). Consequently, observations, interviews, and forced-choice surveys—standard data collection methods in sociology—do not guarantee exclusive access to automatic cognition.

TOWARD BETTER MEASUREMENT OF AUTOMATIC COGNITION

To capture automatic cognition, psychologists have developed measures explicitly designed to prevent respondents from becoming aware of what is being assessed (Petty, Fazio, and Brinol 2009). These are known as implicit measures. Implicit measures attempt to limit deliberate cognition by reducing respondents’ motivation and ability to exert deliberate control over their responses, thus allowing automatic cognition to exert greater influence. Implicit measures also typically require individuals to respond rapidly, which puts additional constraints on deliberate reflection (Gawronski and De Houwer 2014). Implicit measures provide a useful starting point for sociologists interested in measuring automatic cognition. Their effectiveness should not depend on obtaining the right mixture of circumstances, motivations, and capacities, nor on analysts’ ability to identify this mixture in observational settings (Miles 2019). Here, we propose three primary criteria we believe implicit measures must meet to be useful for sociological research. We use these criteria to identify a set of measures to be tested empirically in the next section. The criteria are (1) high scale reliability, (2) weak assumptions, and (3) demonstrated ability to capture automatic cognition.

Scale reliability. This refers to the extent to which an instrument is consistent in its measurement of the underlying construct. Many implicit measures exhibit highly variable scale reliabilities across studies, making them unusable for our purposes (Gawronski and De Houwer 2014:287). Two measures, however, stand out for their typically high reliability (.70 or above): the Implicit Association Test (IAT) and the Affect Misattribution Procedure (AMP). We therefore focus on these measures, both of which encompass several variants (see also Bar-Anan and Nosek 2014; Payne and Lundberg 2014).

IATs are relatively familiar to sociologists. IATs (and variants) require respondents to rapidly sort a series of words or pictures into categories; they measure the strength of association between those categories based on the speed with which items can be correctly sorted (for an overview, see Gawronski and De Houwer 2014). Thus, a positive attitude toward women would manifest as a faster response when women-related items are paired with positive items, and slower responses when paired with negative items. Srivastava and Banaji (2011) demonstrate the utility of the IAT in sociological contexts. Using this measure, they showed that “less conscious” self-views explained collaborative behavior in a mid-sized firm far better than explicit self-report measures. Implicit self-views predicted individuals’ collaborative offers and their likelihood of accepting offers made by others, thereby contributing to the formation of workplace social networks.

AMPs, by contrast, are less well-known in sociology. AMPs rely on priming, or activating representations in memory by exposing respondents to some type of stimuli, often a word or image (known as the prime). In AMPs, respondents are shown a series of primes and neutral images in quick succession. They are instructed to ignore the primes and rapidly rate how pleasant or unpleasant they find the neutral images. Affect generated by the primes are expected to shape respondents’ rating of the neutral images. Unlike the IAT, the AMP does not rely on response times: a positive attitude toward women, for instance, is captured straightforwardly as the level of pleasantness respondents report following women-related primes.

Weak assumptions. The IAT and the AMP differ in their assumptions about how
automatic cognitive processes operate. IAT measures rest on a connectionist model of cognition in which attitudes, identities, and similar constructs are stored in the brain as a network of cognitive elements. A positive attitude toward women might be stored as a connection between the concepts “woman” and “good,” while a traditional frame for gender roles would link “woman” and “home.” The model assumes that activating a node of this network (e.g., through instructions to deliberately think about a specific concept) will automatically activate linked elements, with stronger connections leading to faster activation (Greenwald, McGhee, and Schwartz 1998). IAT measures are based directly on this theoretical framework.

Reliance on a particular (connectionist) model of cognition builds theoretical assumptions into IAT-based measures that could limit their scope of application. To illustrate, an IAT measure of gender attitudes assumes that (1) attitudes are stored as associations between cognitive representations of the attitude object (an image of a woman) and the evaluative constructs (positive and negative words), (2) stronger associations produce more rapid activations of connected elements in the brain, and (3) activation speeds systematically influence response times on the measurement task. Violating any of these assumptions would invalidate the measure. This could occur if individuals are able to sort stimuli at a rapid but steady pace, or if the construct of interest does not rely on associative processing—association, after all, is not a defining feature of automatic cognition (Evans 2008; Stanovich and Toplak 2012).

The AMP rests on weaker assumptions. The procedure assumes the affect generated by the primes spills over into the rating of neutral images through an automatic cognitive process, but it makes no assumptions about the structure of that process (Payne et al. 2005). As a result, the AMP’s validity is not dependent on the accuracy of a particular cognitive model. Importantly, the AMP also allows evaluations to be captured as experienced evaluations (responses to the neutral images), rather than inferred from response times that are assumed to reflect the strength of cognitive associations (Nosek, Hawkins, and Frazier 2011; Payne and Lundberg 2014). The AMP only assumes that respondents are capable of reporting on their internal states, a plausible claim that is routinely made in survey-based research. The AMP thus allows researchers to measure automatically elicited responses—the “outputs” of automatic cognition—without needing a correct model of the way automatic processing operates. In terms of our second criterion, this gives the AMP a clear advantage over the IAT.

**Empirical performance.** To gauge the empirical validity of implicit measures, scholars examine whether they perform in ways consistent with a lack of conscious respondent control. For instance, measures that capture automatic responses should be unaffected by social desirability bias. Several dual-process models also entail that implicit measures should predict spontaneous, unintentional behaviors (Perugini, Richetin, and Zogmaister 2010). Work that examines the IAT for evidence of unintentional responding has yielded mixed results. In a meta-analysis of studies using IAT measures, Greenwald and colleagues (2009) find that measuring socially sensitive topics—such as racial attitudes—reduced the correlation between IATs and explicit forced-choice items (as expected), but it also reduced the IAT’s predictive ability. Furthermore, IATs outperformed explicit measures in predicting outcomes in some (but not all) socially sensitive domains (e.g., race but not sexual orientation). This suggests IATs are less sensitive than explicit measures to motivation to control responses but are still affected by it. IATs also did not predict spontaneous behaviors better than explicit measures (Greenwald et al. 2009; Oswald et al. 2013).

The AMP fares better in this respect. In a recent meta-analysis, Cameron, Brownlannuzzi, and Payne (2012) found that the AMP correlated with explicit measures only under conditions where deliberation was made difficult. Additional studies found that motivations to misreport prejudice and drinking
behavior were unrelated to scores on the AMP, which would be expected if the AMP captures automatic processing (Payne et al. 2013; Payne, Govorun, and Nathan 2008). Importantly, sequential priming measures like the AMP proved equally capable of predicting behaviors regardless of how controllable or socially sensitive they were (Pryor et al. 2013).

Overall, our criteria suggest the AMP may be better suited than the IAT as a general measurement strategy for assessing automatic cognition. However, relatively few studies have examined the extent to which the AMP captures automatic processes, so strong conclusions about its utility cannot yet be drawn. Furthermore, to our knowledge the ability of AMP and IAT to capture automatic cognition has not been compared directly. In what follows, we examine how well IAT and AMP measurement approaches capture automatic cognition in sociologically relevant domains. Rather than gauge automaticity using social desirability or spontaneous behaviors, we use a process dissociation technique that filters out deliberate influences on responses.

PROCESS DISSOCIATION AND PROCESS PURE ESTIMATES

Process dissociation techniques were originally developed in the context of memory research, but these methods have since been used to study social cognition and decision-making, making them well-suited for our purposes (Jacoby, Toth, and Yonelinas 1993; Payne and Bishara 2009). Process dissociation is designed to provide separate estimates of two or more cognitive processes, such as deliberate and automatic influences, on a given behavior. Under the assumption that automatic and deliberate cognitive processes are independent, these estimates are process pure—they each reflect the influence of a single type of cognition (Yonelinas and Jacoby 2012). As such, they provide a benchmark against which other measures of automaticity can be tested. Process dissociation techniques thus allow us to test the extent to which IAT, AMP, and forced-choice measures capture automatic (rather than deliberate) cognition.

Dissociating automatic from deliberate cognition requires observing conditions in which deliberate and automatic influences work in tandem (called an inclusion condition) and conditions in which these influences have opposite implications for behavior (an exclusion condition). For example, Payne, Jacoby, and Lambert (2004) asked respondents to study a list of stereotypically black and white names, each of which were randomly paired with a stereotypically white or black occupation (politician and basketball player, respectively). In a recall task, respondents were then shown each name and asked to select the appropriate occupation for each. If the correct name-occupation combination was stereotype-consistent (e.g., a black name and basketball player), respondents could correctly identify the occupation either because they actually remembered the correct answer or because their memory failed and they relied on an automatic “basketball players are black” stereotype (an inclusion condition). When the name-occupation combination was not stereotype-consistent, respondents could correctly identify the occupation only if they recalled the correct response; reliance on automatic associations would lead to an incorrect answer (an exclusion condition).

Process-pure estimates for each type of influence can be calculated using the probability of giving a particular response in each condition. More formally, the probability that a respondent gives response $R$ in an inclusion condition depends on deliberate, controlled processes ($C$), or if control fails ($1 - C$), on automatically-activated reactions ($A$):

$$P(R|\text{Inclusion}) = C + A(1 - C)$$

In an exclusion condition, the same response only occurs if control fails and automatic processes guide behavior:

$$P(R|\text{Exclusion}) = A(1 - C)$$

$P(R|\text{Inclusion})$ and $P(R|\text{Exclusion})$ can be determined empirically. Solving the equations allows analysts to compute estimates of $C$ and $A$, which can be interpreted as the probabilities that response $R$ is guided,
respectively, by deliberate and automatic influences (Payne and Bishara 2009):

\[ C = P(R|\text{Inclusion}) - P(R|\text{Exclusion}) \]  

(1)

\[ A = P(R|\text{Exclusion}) / (1 - C) \]  

(2)

It is important to note that \( A \) is a measure of all automatic influences on the studied behavior, and not just the ones a researcher might be interested in. As Jacoby and colleagues (1993) note, automatic influences can be divided into a baseline probability of performing a behavior (\( B \)) and automatic effects generated by the study experience (\( P \)), such as a priming effect:

\[ A = B + P \]

The \( B \) component represents any influences on behavior-related automatic processing that arise from stable individual differences, such as working memory, idiosyncratic sensitivity to the study design, and so on. In some cases this might be of interest, but in many applications it represents unwanted noise that obscures the effect produced by the study.

One way to remove the influence of \( B \) from the estimate of interest is to perform the study under two conditions, each of which generates a different automatic reaction. If the same individual completes both versions of the study, \( B \) will be constant across conditions, but \( P \) will vary:

\[ A_1 = B + P_1 \]

\[ A_2 = B + P_2 \]

Subtracting the two estimates of \( A \) will remove \( B \), resulting in a difference score that can be interpreted as the relative difference in the effects of the two study experiences:

\[ A_d = A_1 - A_2 = P_1 - P_2 \]  

(3)

Part A of the online supplement gives further details on the process dissociation technique, including further discussions on interpretation and the method’s robustness to violations of its assumptions. Part B of the online supplement provides an example of process dissociation calculations.

**THE CURRENT STUDIES**

In what follows, we examine how well IAT and AMP measurement approaches capture automatic cognition by determining how well they predict estimates of deliberate and automatic influences on behavior obtained using a process dissociation procedure. We include forced-choice (FC) measures for comparison. To the extent that the IAT and AMP measurement approaches cleanly capture automatic processes, they should accurately predict automatic estimates but have no relation to deliberate cognition estimates. Because FC responses are often produced by a mixture of deliberate and automatic influences, FC items may be related to both deliberate and automatic cognition (Fazio and Olson 2003; Miles 2015).

Although both IAT and AMP measures can be adapted to capture many types of cultural constructs, their largest field of application has been the study of attitudes (Miles 2019). For this reason, we restrict our attention to attitudes, but we highlight other uses of these measures in the discussion. To demonstrate that our results are not specific to any one topic, we compare measures across three domains of sociological interest: politics, morality, and racial attitudes.

**METHODS AND RESULTS**

Data come from three online surveys administered through Amazon’s Mechanical Turk (MTurk) in May 2015, and April to June 2017. MTurk is a widely used online forum allowing researchers to administer studies to an online pool of remunerated respondents. MTurk respondents are not representative of the U.S. population, but in many instances have been shown to respond in ways comparable to traditional research subjects (Mason and Suri 2012; Paolacci and Chandler 2014). All respondents were U.S. residents with a strong record of successful MTurk task
completion (95 percent acceptance rate on previous tasks). We posted our surveys in batches released at different times to capture a greater variety of respondents. The first survey measured political attitudes, the second survey moral attitudes (care and harm), and the third measured racial attitudes. For clarity, we discuss methods and results for each sample separately. Descriptive statistics for all samples are given in Part C of the online supplement.

In place of the traditional IAT, we used a recent adaptation of the IAT called the Brief Implicit Association Test (BIAT; Sriram and Greenwald 2009). The BIAT shares many of the desirable psychometric properties of the IAT (e.g., high reliability) but is shorter and less complex. This makes the BIAT particularly useful to sociologists, for whom study length and respondent fatigue are often concerns (Nosek et al. 2014).

In all samples, process dissociation estimates of automatic and deliberate influences were calculated from responses during survey tasks specifically designed to create inclusion and exclusion conditions while maximizing relationships with the attitudes under study. This allows us to determine whether these measures relate to automatic and deliberate cognition in expected ways, but it sacrifices external validity. The extent to which these measures predict real-world behavior has been addressed (and debated) by others (Cameron et al. 2012; Greenwald et al. 2009; Oswald et al. 2013). Perhaps more importantly, with valid measures in hand sociologists can systematically examine this issue for themselves in the context of their own theories and questions.

**SAMPLE 1: POLITICS**

*Data.* The first survey sampled 397 individuals; of these, we excluded 73 for measurement reasons (described below). We split the remainder of the data into self-reported conservative ($N = 75$) and liberal ($N = 208$) subsamples for analysis (hereafter denoted with the subscripts $C$ and $L$). We outline our reasons for doing so in the Voting Task section. Each subsample was multiply imputed 100 times to deal with cases featuring missing data.

This survey includes only FC measures and an AMP measure. IAT data were collected but rendered unusable by a software error (this was corrected in the next two samples). We administered traditional FC measures after the AMP, and various other items in random order to prevent ordering effects. All items within scales were likewise presented in random order. We adapted each measure to use the same response scale and used the same stimuli for each measure (except where noted below) to reduce differences in measurement approaches that could artificially reduce correlations (Payne, Burkley, and Stokes 2008).

**Measures: forced choice (FC).** Respondents were asked to rate how much they liked or disliked a series of political prompts on a five-point scale from really dislike ($-2$) to really like ($+2$). Prompts included “the Democratic party,” “the Republican party,” “conservative politics (in general),” “liberal politics (in general),” “Barack Obama,” “George W. Bush,” “gay marriage,” and “traditional marriage.” The items were reverse coded as needed and averaged to form a scale with higher values indicating stronger political conservatism ($\alpha_c = .76$; $\alpha_l = .71$).

**Measures: affect misattribution procedure (AMP).** The AMP measures how affect evoked by primes influences respondents’ evaluations of neutral images. Political prime words included the liberal words “progressive,” “left-wing,” “Democrat,” “Barack Obama,” and “gay marriage,” and the conservative words “conservative,” “traditional,” “right-wing,” “Republican,” “George W. Bush,” and “traditional family.” Following standard procedure in AMP research, we used Chinese characters as neutral images. Respondents who reported any reading, writing, or speaking ability in Chinese, Japanese, Korean, or Vietnamese were excluded from the survey. We also warned participants that the words flashed prior to the characters...
might affect their responses, and we instructed them to try to avoid being influenced (Payne et al. 2005). This instruction makes it more likely that systematic influences of the primes on character ratings were unintentional, and hence the result of automatic processes. In each trial, respondents were shown a political prime word for 100 milliseconds (ms), followed by a blank screen for 100 ms, and then a randomly selected Chinese character for 100 ms. The character was then replaced by an image of black-and-white static that remained on the screen until participants responded (see Part D.2 of the online supplement). Respondents were instructed to rate how much they liked or disliked each Chinese character (–2 = really dislike to +2 = really like). The next item was displayed as soon as a response was entered. All prime words were shown twice (in random order), for a total of 24 trials.

The final measure was created by averaging the ratings for liberal-primes trials and conservative-primes trials, then subtracting the former from the latter so that higher values represent more positive attitudes toward political conservatism (Payne, Burkley, et al. 2008; \( \alpha_c = .63; \alpha_l = .67 \)). Subtraction also adjusts for method effects, individual differences in response tendencies, and related validity threats by removing within-individual mean differences in rating tendencies. To guard against non-compliance with instructions, individuals who reported directly rating the prime words instead of the Chinese characters for half the items or more were coded as missing (\( N_c = 13; N_l = 35 \)).

Measures: voting task. The voting task was a multi-trial, rapid-response activity designed to isolate automatic and deliberate influences on voting behavior using process dissociation techniques. During each trial, respondents were shown a randomly selected political prime word (from the set of primes used for the AMP, plus the words “liberal” and “conservative”), followed by a blank screen and then a picture of a potential candidate for the 2016 U.S. presidential election. Each element was displayed for 100 ms. The candidate picture was then replaced by a black-and-white static image until respondents “voted” by pressing one of two keyboard keys, one labeled “would vote for the candidate” and the other “would not vote for the candidate.”

Process dissociation procedures measure automatic and deliberate influences by using differences in response probabilities across inclusion and exclusion conditions. In most studies, these conditions are created by manipulating respondents’ intentions via instructions to complete a task with objectively correct answers (for examples, see Payne and Bishara 2009). An inclusion condition is created when automatic impulses favor giving the correct response; an exclusion condition is created when automatic impulses favor an incorrect response. Because no objectively “correct” voting behavior exists, our study assumes respondents intended to vote in a way consistent with their self-reported political orientation—that is, liberals voting for Democrats, and conservatives for Republicans.13

We measured political orientation using a scale composed of self-rated liberalism/conservatism (–3 = extremely liberal to +3 = extremely conservative) and identification with Democrat/Republican parties (–3 = very strong Democrat to +3 = very strong Republican, \( \alpha = .90 \)). Moderates (defined here as scoring 0 on this scale) could not be classified as liberal or conservative and were excluded from the sample (\( N = 73 \)). To increase the likelihood that respondents would choose candidates based solely on their party affiliations, we used relatively unfamiliar candidates from each political party (as determined in pre-testing) and showed respondents their party affiliations during a learning activity that preceded the voting task.14 In this learning activity, respondents were first asked to study each candidate’s picture, name, and party affiliation. Respondents were then tested on how well they remembered each candidate’s party affiliation. Incorrect answers elicited a repetition of the missed questions. In this way, respondents learned party
information consciously but were unlikely to have strong preexisting automatic associations to candidate images.

For liberals, an inclusion condition was created in the voting task when a liberal prime word was followed by a Democratic candidate. The liberal word primed the respondent with a “vote for” response, and this matched the presumed intention—voting for a Democratic candidate. An exclusion condition was formed when a Republican candidate followed a liberal prime—here the positive reaction generated by the prime opposed intentions to not vote for Republicans. Analogously, inclusion conditions for conservatives paired conservative primes with Republican candidates, and exclusion conditions paired conservative primes with Democratic candidates. Because inclusion and exclusion conditions differed for liberals and conservatives, we divided respondents into liberal and conservative subsamples, and we analyzed each sample separately.

To protect against irrelevant individual differences or method-specific variation that could bias our estimates of $C$ and $A$, we removed baseline differences in the probability of voting for candidates ($B$) using the differencing technique described in Equation 3. We calculated two estimates of $C$ and $A$ for each subsample—one for each set of prime words—and took the difference between the two $A$ estimates. These difference scores ($A_d$) represent the influence of ideologically consistent primes relative to ideologically opposed primes, with larger values representing greater ideological consistency. For the conservative subsample, the score represents the automatic effects of Republican primes relative to the effect of Democratic primes; for the liberal subsample, the score represents the effects of Democratic primes relative to Republican primes. Estimates of $C$ were highly correlated in both subsamples and were combined for analyses ($\alpha_c = .87$; $\alpha_l = .78$). See Part A of the online supplement for further details on the process dissociation methodology.

Sample 1: Results

Table 1 presents coefficient estimates from a series of fully standardized ordinary least squares models that regress deliberate/con- trolling ($C$) and automatic ($A_d$) estimates on FC and AMP measures. Results for the conservative subsample are presented on the left; results for the liberal subsample are presented on the right.

Model 1 indicates that the FC measure has a strong positive relationship with deliberate cognition, with a one standard deviation (SD) change toward conservative political attitudes leading to a .403 SD increase in voting for Republican candidates. In contrast, Model 2 shows that the AMP is not significantly related to deliberate influences on voting ($\beta = .114, p = .408$). Model 3 regresses deliberate cognition on the FC measure and the AMP together and shows that the coefficient for the FC measure remains unchanged, and the estimate for the AMP remains non-significant. We see similar patterns among liberal respondents. Models 4 through 6 indicate that conservative political attitudes measured using FC items have a strong negative relationship with deliberate influences on liberal voting but have no relationship to liberal voting when measured using the AMP. Results in both subsamples thus indicate that FC measures of political attitudes capture deliberate cognition, but as expected the AMP does not.

Estimates from Models 7 through 12 in Table 1 suggest a different pattern for automatic cognition. For conservatives, FC measures are not related to automatic cognition at conventional levels of statistical significance ($\beta = .218, p = .064$), while the AMP has a larger and statistically significant relationship ($\beta = .307, p = .019$, see Models 7 and 8). The estimate for the FC measure is smaller and non-significant when both measures are included in the same model; the AMP coefficient remains large, although it is no longer statistically significant at conventional levels ($\beta = .265, p = .061$, Model 9). This indicates that any contribution the FC items make to
predicting automatic influences on voting behavior can also be obtained using the AMP. Results are similar for liberals, with the AMP outperforming the FC measure. Both measures are correlated with automatic processing, but the estimate for the AMP is twice as large (−.316 versus −.159, Models 10 and 11). When included in the model simultaneously, the FC estimate becomes non-significant, while the AMP coefficient remains essentially the same (β = −.299, \( p < .001 \), Model 12).

Results from Table 1 are consistent with the claim that FC measures capture a mixture of automatic and deliberate cognition, whereas the AMP captures automatic processes alone. Notably, the AMP predicts automatic influences even when FC measures are included in the model, suggesting the AMP captures variation that the FC items do not. Failing to include the AMP would result in a significant loss of predictive power. It also means FC measures cannot serve as a reasonable proxy for the automatic processes captured by the AMP.

**SAMPLE 2: MORALITY**

*Data.* The morality survey was conducted in three waves. The first wave was fielded in April 2017, the second wave about a week later, and the third wave approximately a month following the second wave. We adopted the three-wave design to reduce respondent fatigue and avoid any unanticipated effects from completing back-to-back rapid-response measures. Respondents completed the AMP during the first wave, the BIAT and FC measures during the second wave, and a process dissociation task during the third wave. As in sample 1, we adapted measures to use the same response scales where possible and used the same stimuli for the BIAT and AMP to make measures more comparable. Originally, 361 respondents
participated at wave 1, 311 at wave 2, and 236 at wave three; we multiply imputed missing data ($m = 200$) for a final sample size of $N = 361$ (Enders 2010).\(^{15}\)

**Measures: forced choice (FC).** Respondents were asked to rate how much they liked or disliked a series of prompts related to the care/harm dimension of morality on a scale of really dislike ($-2$) to really like ($+2$). Care prompts included “helping a stranger,” “caring for a sick friend,” “encouraging someone who is disappointed,” “showing compassion,” and “protecting someone from harm.” Harm prompts consisted of “making a cruel remark,” “ignoring a stranger in need,” “laughing at a friend’s misfortune,” “kicking a dog,” and “hurting someone who made you angry” (cf. Graham and Haidt 2012; Graham et al. 2011). Items were reverse coded as needed and averaged to form a scale with higher values indicating a positive attitude toward caring ($\alpha = .83$).

We created a second FC measure by asking respondents to directly rate how much they liked or disliked the images used as primes for the AMP, BIAT, and process dissociation task. To create this second measure, we subtracted average scores for the harm primes ($\alpha = .88$) from average scores for the care primes ($\alpha = .88$). Although using images is not typical for FC measures, this provides a more stringent test for our claims by further reducing differences between the FC and two implicit measures (cf. Payne et al. 2008).

**Measures: affect misattribution procedure (AMP).** Procedures for the AMP were identical to those described previously, except we replaced political prime words with images depicting caring and harmful behaviors (see Part D.3 of the online supplement), and we increased the number of trials to 48. We created the final measure by subtracting the average ratings for harm trials from the average ratings for care trials, so higher values represent more positive attitudes toward caring (Payne, Burkley, et al. 2008; $\alpha = .63$). Individuals who reported directly rating the primes instead of the Chinese characters for half the items or more were coded as missing ($N = 47$).

**Measures: brief implicit association test (BIAT).** The BIAT measures automatic cognition using systematic differences in response speeds during a sorting task. Respondents were shown two categories and instructed to rapidly sort words (or pictures) related to these categories using their keyboard. For each word or picture, respondents pressed the “I” key if it matched either category, or the “E” key for anything else.

Our BIAT began with 16 training trials in which respondents were instructed to sort pictures of animals (mammals/birds) and positive/negative words into the categories “Mammal” and “Things I Like,” respectively. Positive words were “Pleasure,” “Friends,” “Love,” “Health,” “Peace,” “Happiness,” “Joy,” and “Leisure.” Negative words included “Abuse,” “Filth,” “Death,” “Sickness,” “Murder,” “Sadness,” “Suffering,” and “Pain.” This was followed by two blocks of 20 trials each that replaced the category “Mammal” with either “Caring” or “Harmful” (presented in random order) and replaced the images of animals with the same care/harm images as used for the AMP. These blocks were repeated for a total of 80 trials. The first four trials of each block were practice trials designed to familiarize respondents with the new categories; these were excluded from the final measure. Response times for the 64 remaining trials were coded and combined using the recommendations outlined by Nosek and colleagues (2014: Table 8). Higher scores represent more positive evaluations of caring relative to harm ($\alpha = .74$).

**Measures: hiring task.** Like the voting task in the politics sample, the hiring task was a multi-trial, rapid-response activity designed to isolate automatic and deliberate influences on task behavior using process dissociation techniques. Respondents were asked to imagine they had been placed in charge of hiring for a medical company and instructed to hire anyone who had one or more of a given set of degrees. After taking as much time as they
wished to study the list of degrees, respondents completed 12 practice trials at their own pace to consciously learn the correct responses. In each trial, a randomly generated name would appear, followed by one or more degrees (e.g., “Jesse Hughes, MD, PhD”). Respondents chose to hire or not hire applicants using keystrokes (“I” to hire, “E” to not hire). Those who missed more than 20 percent of practice trials were asked to study the list again and given 12 more practice trials.

At that point, respondents were told that for the remainder of the task they were to answer “as fast as you can” so we could gauge how well they make “rapid or ‘snap’ judgments.” Respondents were warned that images would be flashed on the screen prior to applicant names, and they were instructed not to let these images influence their decisions—these images were the same care/harm images from the AMP and BIAT. This was followed by a set of six training trials, then two rounds of 32 trials each. Each trial consisted of a randomly selected harm/care prime displayed for 100 ms, followed by a blank screen for 100 ms, and then a randomly selected applicant name and degree combination for 100 ms. The applicant name was then replaced by a static image until a response was entered, at which point the next trial immediately began.

We expected that seeing an image of caring behavior would generate a positive affective reaction that would predispose respondents to hire the applicant who followed it. Thus, an inclusion condition was created when a care prime was followed by an applicant possessing one of the hirable degrees, and an exclusion condition arose when a care prime preceded a non-hirable degree. We used probabilities of giving a “hire” response in both conditions to calculate C and A. We also calculated C and A using harm-prime trials and subtracted the A estimate from the care-prime A estimates to control for individual and method effects. These difference scores (A_d) represent the influence of care primes relative to harm primes. As in the politics sample, we combined C estimates into a single score (α = .82).

**Sample 2: Results**

We turn first to the results for deliberate cognition. In contrast to the politics sample, the hiring task has objectively correct responses that do not depend on respondents’ preferences. This means C captures non-evaluative deliberate processes (e.g., working memory), and so should be unrelated to all attitude measures, whether explicit or implicit. The results in Table 2 generally support this claim—coefficients for the traditional FC, picture-based FC, and AMP are all close to zero. However, the coefficient for the BIAT is unexpectedly much larger and statistically significant (β = .193, p = .006 in Model 5). This might be because the same types of non-evaluative cognitive processes required to provide correct responses during the hiring task might also be needed to accurately sort items during the BIAT. If this is true, the BIAT captures deliberate processing but not necessarily processing related to the topic under study (morality).

The pattern of results is different for automatic cognition. Models 6 and 7 of Table 2 show that neither FC measure predicts automatic influences on task performance. This might be because the FC measures do not capture automatic processes, or because they measure different automatic processes from those elicited by the images used during the hiring task (e.g., a sense for what is socially approved, as opposed to personal evaluative reactions). Surprisingly, the BIAT also fails to predict automatic processing (Model 9). The AMP measure, however, predicts automatic influences, both when modeled as the sole predictor and when included alongside the other variables (Models 8 and 10). These results are consistent with the claim that FC measures do not necessarily capture automatic processes, but the AMP does.

**SAMPLE 3: RACIAL ATTITUDES**

**Data.** The racial attitudes survey mirrors the design of the morality survey in every respect, but the questions and prime materials were
modified to address racial attitudes rather than morality. Originally, 330 respondents participated at wave 1, 283 at wave 2, and 229 at wave three. Missing data were multiply imputed ($m = 200$) for a final sample size of $N = 330$.

**Measures: forced choice (FC).** Respondents were asked to rate how much they liked or disliked a series of prompts related to racial attitudes on a scale of really dislike ($-2$) to really like ($+2$). We used prompts similar to those common in literature on racial attitudes (Sears, Sidanius, and Bobo 1999), such as “welfare,” “affirmative action,” “voter ID requirements,” “Black Lives Matter movement,” “immigration restriction,” and “food stamps.” The items were averaged to form a scale with higher values indicating a positive attitude toward the prompts (reverse coded where needed, $\alpha = .84$).

Respondents also directly rated how much they liked or disliked facial pictures of black and white individuals that were used as prime images for the implicit measures and process dissociation task, using the same scale as the previous FC measure. We subtracted the average ratings for the white primes ($\alpha = .87$) from the average rating for black primes ($\alpha = .88$) to create the final scale, where higher scores represent greater liking for black relative to white primes.

**Measures: affect misattribution procedure (AMP).** Procedures and number of trials for the AMP were identical to those described for sample two, however here the

<table>
<thead>
<tr>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Est/SE</td>
<td>$\beta$</td>
<td>Est/SE</td>
<td>$\beta$</td>
<td>Est/SE</td>
</tr>
<tr>
<td>FC</td>
<td>0.024</td>
<td>(0.035)</td>
<td>0.050</td>
<td></td>
</tr>
<tr>
<td>FC (picture)</td>
<td>0.041</td>
<td>(0.022)</td>
<td>0.134</td>
<td></td>
</tr>
<tr>
<td>AMP</td>
<td>-0.009</td>
<td>(0.044)</td>
<td>-0.014</td>
<td></td>
</tr>
<tr>
<td>BIAT</td>
<td>0.095</td>
<td>(0.033)</td>
<td>0.201**</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model 6</th>
<th>Model 7</th>
<th>Model 8</th>
<th>Model 9</th>
<th>Model 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Est/SE</td>
<td>$\beta$</td>
<td>Est/SE</td>
<td>$\beta$</td>
<td>Est/SE</td>
</tr>
<tr>
<td>FC</td>
<td>-0.009</td>
<td>(0.031)</td>
<td>-0.020</td>
<td></td>
</tr>
<tr>
<td>FC (picture)</td>
<td>-0.002</td>
<td>(0.020)</td>
<td>-0.007</td>
<td></td>
</tr>
<tr>
<td>AMP</td>
<td>0.082</td>
<td>(0.040)</td>
<td>0.152*</td>
<td></td>
</tr>
<tr>
<td>BIAT</td>
<td>0.011</td>
<td>(0.029)</td>
<td>0.027</td>
<td></td>
</tr>
</tbody>
</table>

$N = 361$

**Note:** Standard errors are in parentheses; $\beta$ column presents fully standardized coefficients; intercepts not shown. FC = forced choice, AMP = affect misattribution procedure.

*p < .05; **p < .01; ***p < .001 (two-tailed tests).
primes consisted of facial pictures of black and white men (see Part D.4 of the online supplement). The final measure was created by subtracting the average ratings for white-prompt trials from the average ratings for black-prompt trials, so that higher values represent more positive attitudes toward blacks than toward whites ($\alpha = .69$). Individuals who reported directly rating the prime words instead of the Chinese characters for half the items or more were coded as missing ($N = 31$).

Measures: brief implicit association test (BIAT). Procedures for the BIAT mirrored those used in the morality sample, except that pictures of black and white men replaced morality primes. The BIAT is coded such that higher scores represent more positive feelings toward blacks relative to whites ($\alpha = .45$).

Measures: hiring task. Respondents in the race sample completed a hiring task identical in design to that described for the morality sample. The only difference is that morality primes were replaced by the black/white faces used for the AMP and BIAT. We calculated $C$ and $A$ using the inclusion and exclusion probabilities of giving a “hire” response following both black and white trials, then subtracted the latter estimates of $A$ from the former to adjust for individual and method effects. These difference scores ($A_j$) represent the influence of black primes relative to white primes. As in the politics and morality samples, we combined $C$ estimates into a single score ($\alpha = .85$).

Sample 3: Results
As in the morality sample, deliberate influences on performance during the hiring task depend on non-evaluative cognitive processes, and consequently should be unrelated to any of our racial attitude measures. Results in Table 3 support this contention. Neither FC measure, nor the AMP or BIAT, predict $C$. The pattern of results for automatic cognition is different. Here, ratings of prime images have a strong relationship with automatic cognition that persists even when adjusting for other measures (Model 10). No other measures are related to automatic processing except the AMP, which shows an effect about as large as observed in the morality sample ($\beta = .143$, $p = .039$, Model 8). However, this effect disappears when the other variables are added in Model 10, due almost entirely to the AMP’s correlation with the picture-based FC measure ($r = .35$). Results once again suggest the AMP captures automatic cognition, but they also indicate that direct ratings of images can do so as well, and they might have particular utility in studying racial attitudes.

DISCUSSION
Summary of the Argument and Results
Sociologists working across a wide variety of substantive subfields rely on automatic cognition to theorize important social processes. However, the measurement of automatic processes remains underdeveloped in sociology. Scholars often rely on ad hoc measures that require strong assumptions to detect automatic cognition. This hinders empirical applications of dual-process models, impedes theory testing and development, and complicates efforts to synthesize existing empirical work into a systematic body of knowledge. In this article, we argue that progress in understanding the cognitive bases of social phenomena will be facilitated by the use of standard measurement practices specifically designed to capture automatic cognition. We proposed three criteria for identifying useful measures: high scale reliability, weak assumptions about the underlying cognitive processes, and a demonstrated ability to capture automatic processes. We tested the validity of two implicit measures—the AMP and the BIAT—by comparing their ability to predict process-pure estimates of automatic cognition alongside traditional forced-choice (FC)
measures in three samples and across three substantive domains.

Our results strongly favor the AMP over FC and the BIAT as a valid measure of automatic processes useful for sociologists. As Table 4 demonstrates, the AMP was the only measure that consistently predicted automatic influences on behavior. The fact that its performance was consistent across all three samples suggests its ability to capture automatic cognition is not limited to particular cultural domains. The evidence in favor of the AMP is particularly strong given that in the morality and racial attitudes studies, the AMP was administered more than a month prior to the process dissociation tasks. Furthermore, the AMP measured automatic processes cleanly, as evidenced by its universal failure to predict deliberate cognition (Table 4).

This stands in contrast to traditional FC measures, which we found to be strongly tied to both automatic and deliberate cognition in the politics sample, supporting prior claims that FC measures capture a mixture of both processes (Fazio and Olson 2003; Miles 2019). Asking respondents to directly rate prime pictures improved the performance of FC measures in the race sample, where the pictured-based FC measure was by far the strongest predictor of automatic influences on task behavior. This pattern did not replicate in the morality sample, so it is unclear whether the effect is due to chance, or because rating images taps automatic evaluations better in some domains than others. The latter interpretation is consistent with theories that emphasize the role of visual cues in activating cultural frames (e.g., Ridgeway and Kricheli-Katz 2013).

Table 3. Linear Regression of Deliberate and Automatic Influences on Hiring Task Behavior for FC, AMP, and BIAT Measures (Race)

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6</th>
<th>Model 7</th>
<th>Model 8</th>
<th>Model 9</th>
<th>Model 10</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est/SE</td>
<td>β</td>
<td>Est/SE</td>
<td>β</td>
<td>Est/SE</td>
<td>β</td>
<td>Est/SE</td>
<td>β</td>
<td>Est/SE</td>
<td>β</td>
</tr>
<tr>
<td>FC</td>
<td>.033</td>
<td>.125</td>
<td>.034</td>
<td>.098</td>
<td>−.004</td>
<td>−.008</td>
<td>.003</td>
<td>.004</td>
<td>−.015</td>
<td>−.022</td>
</tr>
<tr>
<td></td>
<td>(.017)</td>
<td></td>
<td>(.025)</td>
<td></td>
<td>(.039)</td>
<td></td>
<td>(.046)</td>
<td></td>
<td>(.047)</td>
<td></td>
</tr>
<tr>
<td>FC (picture)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>−.006</td>
<td>−.026</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.018)</td>
<td></td>
<td>(.025)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMP</td>
<td>−.004</td>
<td>−.008</td>
<td>.066</td>
<td>.143*</td>
<td>.052</td>
<td>.086</td>
<td>.029</td>
<td>.047</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.039)</td>
<td></td>
<td>(.032)</td>
<td></td>
<td>(.040)</td>
<td></td>
<td>(.040)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIAT</td>
<td>.066</td>
<td>.143*</td>
<td>.052</td>
<td>.086</td>
<td>.029</td>
<td>.047</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(.039)</td>
<td></td>
<td>(.035)</td>
<td></td>
<td>(.040)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>361</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Standard errors are in parentheses; β column presents fully standardized coefficients; intercepts not shown. FC = forced choice, AMP = affect misattribution procedure. *p < .05; **p < .01; ***p < .001 (two-tailed tests).
The BIAT did not predict automatic influences but did unexpectedly predict deliberate cognition in the morality study. We hypothesized that the BIAT and the process dissociation task might require similar types of intentional cognition to complete, which would explain the association. This interpretation remains tentative, however, given that the effect did not replicate in the racial attitudes study. The inability of the BIAT to predict automatic cognition in this study is consistent with prior work demonstrating the IAT does not always relate to other measures in theoretically expected ways (e.g., Oswald et al. 2013).

One possible reason for this is that IAT-type measures like the BIAT rely on a number of untested assumptions about what attitudes are and how they are processed. These assumptions might not hold in all contexts, compromising the measure and inhibiting its performance. Alternatively, the BIAT’s poor performance might also be due to its unique design features, in which case other IAT-type measures might perform better, although this would be inconsistent with prior work showing that the BIAT and IAT behave similarly (Bar-Anan and Nosek 2014). A final possibility is that the BIAT simply captured a form of automatic cognition that was not related to the automatic processes elicited during the survey-based tasks; if true, the BIAT might perform better in other settings.

Reliance on samples of online respondents may raise validity concerns about our study. Social scientists have increasingly turned to Mechanical Turk as a source of rapid, inexpensive survey data, but the pool of respondents actively completing studies at any given time is relatively small. As a result, MTurk workers tend to be experienced research participants (Chandler, Mueller, and Paolacci 2014; Stewart et al. 2015). In the current study, close to 70 percent of respondents in the morality and racial attitudes studies reported having completed more than 30 studies in the month prior to beginning the first wave of the survey. This could lead to a high level of familiarity with forced-choice scales and popular psychological measures like the IAT. Familiarity might have allowed respondents to rely more heavily on automatic processes when completing FC items, which means this study might overestimate the ability of FC measures to capture automatic cognition. It also offers another explanation for the poor performance of the BIAT: effect sizes tend to shrink as familiarity with the IAT increases (Nosek, Greenwald, and Banaji 2007), possibly because it compresses the differences in response latencies on which the BIAT relies.17 The extent to which familiarity explains the BIAT results relates to its usefulness as a measure, suggesting it might only be useful in inexperienced populations.


\[
\text{Percent of tests consistent with automaticity} \quad 50\% \quad 25\% \quad 100\% \quad 50\% \quad 100\% \quad 100\% \quad 50\% \quad 0\%
\]

Note: X = effect detected in either bivariate or multivariate analysis, blanks = no effect; dashes = tests that were not performed. Automaticity implies the absence of an effect for C, and the presence of an effect for \( A_d \).
Against this backdrop, the consistent performance of the AMP is particularly impressive and suggests it might be a more widely applicable—and hence more useful—measure.

**Implications for Sociological Practice**

Our work has a number of implications for sociological practice. First, it recommends against using FC scales as measures of automatic cognition. These measures probably do tap automatic processes to some degree, but they also risk being mixed together with deliberate cognition unless the right individual and contextual conditions prevail at the time of measurement, which can be difficult to establish. Using FC items forces researchers to rely on analytic inferences from ambiguous data, rather than more reliable empirical evidence of automatic processing. Importantly, results from our politics study also indicate that FC measures cannot always be assumed to only tap deliberate cognition. Using FC items to measure deliberate cognition thus may only be possible if there are strong theoretical reasons to assume it does so cleanly, or when measures of automatic cognition are included alongside FC items to control for shared variance (e.g., Models 3 and 6 in Table 1). An important implication of this study is that FC items provide clean measures of neither type of cognition.

Second, this study provides strong reasons to incorporate the AMP into the corpus of sociological measurement tools. AMPs generally focus on determining respondents’ evaluations of people, groups, or ideas. All of these have a long history in the sociological study of symbolic and social boundaries and broader evaluation processes (e.g., Lamont 2012; Lamont and Molnár 2002), suggesting AMPs can make a tangible contribution to many areas of social research. Research on cultural tastes, for instance, shows that preferences map in different ways onto social structural positions. Bourdieu’s work suggests these distinctions should be largely automatic. However, a heavy reliance on interview, observational, and forced-choice survey methods makes it difficult to know if this is the case, thus creating a crucial gap between what is theorized and what is known (Bourdieu 1984; Goldberg 2011; Peterson and Kern 1996). To address this, a cultural tastes AMP could be constructed by using images of cultural objects (e.g., artwork, band logos), short sound clips, or even smells as primes, and observing their effect on evaluations of neutral images. This measure could then be used to help determine if cultural tastes are rooted in automatic preferences, or are tied to some other process, such as deliberate efforts at self-presentation.

Recent evidence also indicates that semantic content can be misattributed in the same way as affect, allowing the AMP to be adapted to measure many types of meaning (Imhoff et al. 2011; Sava et al. 2012). Sava and colleagues (2012), for example, measured automatic self-concepts by priming respondents with words associated with personality traits, and then asking them to rate whether the Chinese characters that followed “fit” their personality. Similarly, AMP-like measures could be designed to capture any number of constructs of interest to sociologists, including beliefs, identities, stereotypes, and other cultural associations (see Miles 2019). To give just two examples, scholars interested in moral boundaries might use images of different social groups as primes and ask respondents to guess whether neutral images that follow them have an immoral, neutral, or moral meaning (Lamont 1992; Lamont and Molnár 2002; Miles 2014b). Or gender researchers could capture automatic gender beliefs by priming respondents with images of men and women, and then asking them to guess whether unfamiliar characters have meanings associated with home or work (Coltrane 2000; Davis and Greenstein 2009), represent STEM fields or humanities (Nosek et al. 2009), or express stereotypically masculine or feminine personality traits (Auster and Ohm 2000).18

A third implication is that scholars can begin formally testing theories that rely on automatic cognition. For example, claims about the effects of automatically activated
cultural frames on interaction, implicit prejudice on police action and legal proceedings, and *habitus* on educational achievement can be more directly assessed with measures of automatic cognition (Davenport et al. 2011; Gaddis 2013; King and Johnson 2016; Ridgeway and Krcheli-Katz 2013). Similarly, scholars can begin to address questions about how and when cultural elements are deployed, how automatically processed elements interact with each other, and how and to what extent they interact with deliberate processes (e.g., Vaisey 2009; Vila-Henninger 2015). With standard measures of automatic cognition, researchers can test competing models and refine theoretical claims.

Finally, having valid measures of automatic cognition can advance research on a variety of substantive issues. For example, Lizardo (2017) argued that a dual-process understanding of cultural learning and deployment can unravel seeming paradoxes in research on educational attainment, class-based socialization, and elite cultures, in which what people say seems unrelated to what they do. Lizardo proposed that these puzzles resolve themselves once we recognize that different forms of culture are learned and deployed in different ways and need not overlap or be consistent with one another. This solution applies empirical results from outside of sociology to provide a plausible interpretation of data in several areas of sociological inquiry. With measures of automatic cognition, this interpretation can be empirically tested by examining whether declarative and nondeclarative cultural forms correspond, and which relate to behavior. Such measures could also be used to examine the social origins and distributions of schemas and other cultural constructs, a topic that has commanded attention in recent years (Boutyline 2017; Goldberg 2011; Hunzaker 2017; Miles 2014b). Traditionally, efforts to measure cultural schemas have relied on large numbers of forced-choice items. Using the AMP or other valid implicit measures would better capture the automatic nature of schemas (Miles 2019; for an example, see Hunzaker 2017).

### Challenges and Suggestions for Future Research

The AMP provides a way to capture many of the cultural constructs discussed in existing theories, but fully testing these models will require using (or developing) measures of action schemas, skills, and other cultural constructs that predict action (see Miles 2019). This underscores an important point: however useful the AMP is, it should be seen as a starting point and one tool among (potentially) many. Scholars should explore the validity and utility of other implicit measures and, as needed, develop measures of their own (Bar-Anan and Nosek 2014; Gawronski and De Houwer 2014; Miles 2019). We believe the criteria outlined above of high scale reliability, weak assumptions about the underlying cognitive processes, and a demonstrated ability to capture automatic processes offer useful practical guidelines going forward. Other criteria might be useful in particular applications, such as insensitivity to extreme scores or outliers, criteria on which the IAT and BIAT may have an advantage (for a discussion, see Bar-Anan and Nosek 2014; Payne and Lundberg 2014).

Using appropriate measures of automatic cognition requires collecting data in ways that are unfamiliar to many sociologists. There are many accessible guides to using the AMP and various IAT measures (including the BIAT), but we recommend starting with a recent piece by Miles (2019) that not only outlines how the measures work, but also gives recommendations for how they might be adapted to capture the constructs of interest to sociologists. Once this foundational knowledge is in place, the details of actually administering the measure can be simplified by using existing software. Implicit tests like the BIAT and AMP can be implemented using commercial packages like Inquisit, or free programs like FreeIAT or WebIAT. Both classes of products provide options for administering implicit tests online, so researchers are not restricted to local samples. With some programming experience, researchers can even create these measures.
themselves. The BIAT, AMP, and process dissociation tasks used in this study, for instance, were created using javascript coding in Qualtrics, a popular online survey platform.

Dual-process models have already made significant contributions to sociological theory, but these theoretical innovations need to be matched with methodological advances so that future theory building can be based on a solid empirical foundation. We demonstrated that measuring automatic cognition can be accomplished using the AMP, and we offered criteria that can be used for assessing the utility of other measures of automatic cognition. Capturing automatic cognition will necessarily entail adopting new strategies for collecting data, but the practical difficulties this raises are manageable, especially when considering the potential payoff of valid, reliable, and interpretable measures. With these measures in hand, sociologists will be able to move beyond the question of whether they are capturing automatic cognition to substantive questions about how automatic processes develop and operate in the context of a complex social landscape.

Acknowledgments and Data Note
We would like to thank Stephen Vaisey, Vanina Leschziner, and several anonymous reviewers for helpful comments on an earlier version of this manuscript. Data and code files to replicate analyses are posted on the first author’s website (www.andrewamiles.com).

Funding
This research was supported by grants from the sociology departments at Duke University and the University of Toronto.

Notes
1. The utility of deliberate cognition is also emphasized by agency researchers who argue for the importance of “reflexivity”—a process of deliberate internal conversation and reasoning—in producing human action (Archer 2013; Elder-Vass 2007; Hitlin and Johnson 2015).
2. Some scholars add that implicit measures should capture a construct that was produced in an automatic fashion (Gawronski and De Houwer 2014).
3. Most implicit measures can be used to capture constructs that contribute to “culture in thinking” processes, which are processes that influence action but do not directly produce it (Lizardo et al. 2016). Measures of “culture in action” processes (e.g., habits) are rarer, and so are not considered further here (but see Miles 2019).
4. Given our interest in social action, one could argue that the measure should have a demonstrated ability to predict behavior. We believe, however, that the link between implicit cognitive processes and behavior should be determined empirically, rather than built into the measurement technique as an assumption.
5. Test-retest reliabilities for the IAT and the AMP range from approximately $\alpha = .30$ to $\alpha = .70$ (Bar-Anan and Nosek 2014; Gawronski et al. 2017).
6. See Payne and Lundberg (2014) for a review of evidence that supports the misattribution assumption.
7. Independence means that altering one type of cognition does not alter the other. This assumption does not hold in all situations, but Yonelinas and Jacoby (2012) provide evidence that process dissociation tasks meet the independence assumption. For example, divided attention, age, and rapid response affect deliberate but not automatic estimates, consistent with theoretical expectations that these variables influence deliberate cognition alone. Furthermore, automatic estimates obtained using process dissociation are typically very similar to those obtained using other methods thought to provide process-pure estimates.
8. Another option for removing $B$ is to directly measure $B$ and then subtract it from $A$, giving $B + P - B = P$. A measure of $B$ can be obtained by observing the rate at which respondents perform the behavior ($R$) when the effects of the study design are minimal. In the race/occupation study cited in the text, this might be achieved by asking respondents to guess the occupations associated with each name without allowing them the opportunity to study the list of correct answers, leaving them only their intuitions to rely on.
9. All tasks required respondents to perform behaviors that are likely influenced by automatic attitudes. Specifically, we assume that the positive or negative evaluative reaction generated by seeing primes related to politics, morality, and race will influence whether respondents perform positive actions, operationalized in our study as voting for a candidate or hiring an applicant. Because all measures are designed to capture the same attitudes, all have the same potential to predict task behavior. Variation in actual predictive power can thus be attributed to differences in each measure’s ability to capture automatic cognition.
10. In particular, response times did not record accurately for the IAT measure. This error did not affect respondents’ experience while taking the survey, but it prevented a planned comparison between the IAT and AMP.
11. Using words as primes instead of images made it easier to evoke liberalism and conservatism in a generic sense and reduced reliance on specific issues or individuals associated with these ideologies. It also avoided any potential complications of using images of political figures as both primes and rating targets during the voting task. Samples 2 and 3 used images because care, harm, and race are more easily visualized.

12. We strongly suggest that sociologists interested in using the AMP develop and validate alternative neutral images. Reliance on Chinese characters forces researchers to exclude respondents familiar with these characters from their research sample, which limits the AMP’s application. Moreover, in studies investigating attitudes on immigration, Asian Americans, foreign policy, and other topics, Chinese characters are not neutral images and could bias results.

13. Consistent violations of the assumption that respondents intended to vote in ideologically consistent ways (e.g., a liberal intending to vote for Republicans) would alter the inclusion and exclusion conditions and change the scale of C and A but would not bias estimates—the measure of C would accurately capture the level of deliberate control, but the sign would be reversed (see Part A of the online supplement). A lack of intention would likewise create no bias—voting decisions would be driven entirely by automatic processes, and the estimate of C would be close to 0. Bias would only occur if respondents intended to vote for a mixture of Republicans and Democrats, thus creating no consistent inclusion and exclusion conditions. This scenario seems unlikely given that candidates were relatively unfamiliar, and therefore respondents had few (if any) reasons to prefer particular candidates other than because of their party identification.

14. We also included several recognizable candidates to increase the face validity of the task. Our results were substantively similar with or without these candidates, so we opted to retain them for analyses. All candidates are listed in Part D.1, Table S3, in the online supplement.

15. We used separate imputations to calculate scale reliability coefficients and for analyses. Calculating reliabilities requires imputing individual scale items, but this approach can introduce considerable noise into analyses because it allows chance correlations associated with individual scale items to play a greater role in determining imputed values. This is particularly problematic if individual scale items have modest correlations with key study variables, as is the case in this study. For this reason, analysis imputations included complete scales rather than individual scale items. Note, too, that the loss of respondents between waves necessarily reduces efficiency and may introduce biases. Multiple imputation improves efficiency by recovering some of the lost information, and it can produce unbiased estimates under weaker assumptions than listwise deletion, which is the usual alternative (Enders 2010).

16. The set of given names was chosen to be gender neutral, and surnames were chosen from a collection of common surnames. See Part D.5 of the online supplement for more information.

17. The high level of familiarity is suggested by an email received by the first author, in which a respondent noted that “[w]e here on MTurk are basically professional IAT-takers.”

18. Future work might also use AMP-like measures to test the validity of claims that skillful analysis of qualitative data can reveal automatic cognition. For example, do the (assumed) implicit gender schemes uncovered during an interview match those detected using the AMP?

References


Andrew Miles is an Assistant Professor of Sociology at the University of Toronto. His work lies at the intersection of the sociology of culture, social psychology, and moral and cognitive psychology. His research focuses on the social development of different moral cultures, the role moral constructs play in predicting behavior (e.g., values, moral identities), and the effects of moral behavior on emotions. He also studies how cognitive processes affect action and has an abiding love for learning and teaching quantitative methods.

Raphaël Charron-Chénier is an Assistant Professor of Justice and Social Inquiry in the School of Social Transformation and an Affiliated Professor of Sociology in the School of Social and Family Dynamics at Arizona State University. His research focuses on racial inequality in the context of consumer and credit markets. His recent work examines the racialization of payday lending through processes of predatory inclusion.

Cyrus Schleifer is an Assistant Professor of Sociology at the University of Oklahoma. His research examines different statistical approaches to modeling change in religious belief, practice, and belonging. He is currently focusing on modeling gender inequalities within religious institutions, uncovering the changing patterns in the relationship between religion and higher education, and mapping national-level changes in religious belonging within the United States.