4 Implicit Measures in the Experimental Psychology of Religion

Jonathan Jong

EXPLICIT MEASURES AND THEIR DISCONTENTS

Pace the radical behaviourists, psychologists are interested in mental states. We want to know how people are feeling and what they are thinking. Often, our interests are more specific: we want to know how people feel about certain things, and what they think about certain things. We want to know how some stimulus makes them feel; when they are more likely to believe what someone is telling them; what kinds of faces they like. And of course, psychologists who study religion are interested in such things too, in a specific context. We’re interested in how people feel when they pray; what influences their beliefs in supernatural agents; their attitudes toward their gods. Above and beyond religious behaviours, we are interested in religious cognition and emotions. But those of us interested in such things are immediately faced with the problem of measurement. How can we access (and assess) what people are feeling and thinking?

The commonsensical answer to this question is, of course, to ask them. And that’s just what social psychologists and social scientists more broadly have done for decades. The vast majority of our data come from self-report measures of various kinds, both qualitative and quantitative. Much of our qualitative data come from ethnographic work undertaken by cultural anthropologists, but face-to-face interviews (including clinical interviews) and textual or discourse analyses are also used (see Hood, Spilka, and Gorsuch 2009 for a recent survey of research methodology in the psychology of religion). The quantitative data, on the other hand, are mostly collected via questionnaires, of which over a hundred are currently available (Hill and Hood 1999). Despite the preponderance of questionnaires, researchers still sometimes find the need to design their own self-report questionnaires for specific purposes (e.g., Cohen, Shariff, and Hill 2008; Gibson 2005; Jong, Blumke, and Halberstadt: submitted). Part of the problem stems from the piecemeal approach taken by CSR, which focuses on specific aspects of religion (e.g., belief in supernatural agents), rather than assessing general “religiosity/religiousness” (e.g., Rohrbaugh and Jesson 1975), types of religiosity (e.g., Altemeyer and Hunsberger 1992) or general attitudes toward
religion and spirituality (e.g., Bardis 1961). Furthermore, many of the well-established questionnaires were designed for Western, Christian participants and are ill-suited in more non-Christian or pluralistic contexts (e.g., Christian Orthodoxy Scale; Fullerton and Hunsberger 1982).

So, we’re still doing this. Designing, testing, using, and interpreting questionnaires; these are basic skills in a social psychologist’s repertoire, and that’s not going to change any time soon. Questionnaires are very useful things. However, we’ve also long acknowledged the limitations of such direct or overt or explicit measures of psychological variables.

There are, broadly speaking, two classes of problems with explicit measures of psychological variables, as revealed in the social psychological literature on attitudes. The first is the “strategic responding” problem (Wittenbrink and Schwarz 2007: 2): participants might not always be honest when asked directly about their feelings, beliefs, and desires. For example, participants are more likely to report more racially prejudiced attitudes under conditions of increased anonymity or privacy (e.g., Plant, Devine, and Brazy 2003). The standard interpretation of this trend is that people are unwilling to report their true feelings, beliefs, and desires if they’re deemed socially undesirable. Besides being vulnerable to the social desirability effect, explicit measures may also serve as demand characteristics: cues which lead participants to form beliefs about the experimenters’ expectations (Orne and Whitehouse 2000). As all budding experimental psychologists are warned, demand characteristics can influence participants’ responses in undesirable (and even unpredictable) ways, thereby extraneously affecting the results of the experiment.

The second class of problems arises when we consider that people might not be aware of their own feelings, beliefs, and desires, or at least some aspects thereof. Even if explicit measures can accurately capture explicit attitudes, might there not also be implicit attitudes? Over the last two decades, this view—that some attitudes are held or formed automatically and even unconsciously—has established itself as social cognitive orthodoxy. The literature is now replete with dual-process models of cognition, which distinguish between the implicit and explicit (e.g., Nosek 2007), or the automatic and controlled (e.g., Bargh and Chartrand 1999), or the unconscious and conscious (e.g., Dijsterhuis and Nordgren 2006), or the heuristic and systematic (e.g., Chen and Chaiken 1999). Although there’s still much empirical and theoretical work to be done here, it’s now clear to social cognitive psychologists that traditional self-report measures are inadequate to capture all that we’re interested in.

Of course, these methodological concerns apply to research in the cognitive science and psychology of religion too. The tendency towards “theologically correct” (Barrett and Keil 1996) responding is explicable both in terms of strategic responding and of dual-process models of cognition. People might be motivated to report lower or higher levels of belief in different contexts. For example, fundamentalist Christians and militant atheists
might similarly be reluctant to admit being swayed by an experimental manipulation (e.g., mortality salience priming). They might therefore still parrot theologically (or atheologically) correct responses, despite their newfound doubts. On the other hand, Barrett and Keil (1996) demonstrated that religious believers simultaneously claimed to hold theologically correct beliefs, while also using more anthropomorphic God concepts in a reading comprehension task. Perhaps—as Barrett and Keil (1996) and others have suggested—theologically correct beliefs are held explicitly, while anthropomorphic (i.e., theologically incorrect) beliefs are held implicitly. Either way—whether theologically correct responding is an example of strategic responding, or a product of dual-process cognition, or both—traditional self-report measures are limited in their abilities to capture people’s actual and/or implicit religious beliefs.

**IMPLICIT MEASURES, PART 1: “LOW-TECH” OPTIONS**

In view of these concerns over the limitations of explicit measures, psychologists have devised many *indirect* or *covert* or *implicit* measures. The earlier techniques were direct attempts to reduce strategic responding. Indeed, they also revealed the extent of socially desirable responding. For example, Sigall and Page (1971) found that white American participants ascribed more negative traits to African Americans when they were connected to a bogus lie detector. That is, when participants believed that the experimenters had a pipeline to their minds, they were more willing to express socially undesirable, racist attitudes. Less interestingly, social psychologists have long known to increase assurances of anonymity and to avoid face-to-face interviewing, as these are effective ways to reduce strategic responding (Nederhof 1984). Indeed, in many social psychology laboratories, participants are assigned subject numbers, and seated in private, light- and sound-attenuated booths as they complete their tasks (e.g., questionnaires).

1 Partially Structured Measures

Besides such supplements to more or less traditional questionnaires, psychologists have also developed a variety of partially-structured measures (Cook and Sellitz 1964): measures involving participants’ interpretations of stimuli (e.g., pictures, narratives). It must be said that these measures are predominantly used by clinicians and researchers from the psychoanalytic tradition, in the form of projective techniques like the famous Rorschach inkblot test (Rorschach 1927), and the less famous (but probably more widely used) Thematic Apperception Test (Murray 1943). In both these tests, participants’ interpretations of visual patterns or drawings of events are interpreted to infer personality characteristics and other psychological variables. In these forms, partially structured measures have not fared well.
among experimental social psychologists, largely for their poor psychometric properties; the research of the reliability and validity of these measures have not be reassuring (Vargas, von Hippel, and Petty 2004).

However, there have been recent attempts to revive and reform partially-structured measures in social psychology. For example, Vargas, von Hippel, and Petty (2004) have made a concerted effort to design and test a narrative-based, partially structured measure. Consider the following vignette:

Mary didn’t go to church once the whole time she was in college but she claimed that she was still a very religious person. She said that she prayed occasionally and that she believed in Christian ideals. Sometimes she watched religious programs on TV like the 700 Club or the Billy Graham Crusade. (Vargas et al. 2004: 197)

Vargas et al. (2004: Study 4) presented participants with 20 of such vignettes, and asked them to respond to two questions on an 11-point scale (from “Not at all religious” to “Extremely religious”): “How religious was the behaviour Mary performed?” and “How religious do you think Mary is, in general?.” As hypothesized, they found that participants’ ratings on these questions were significantly correlated with self-report measures of religious attitudes and religious behaviours. Indeed, they found that considering both participants’ ratings in the partially-structured measure and in the self-report religious attitudes questionnaire provided better predictions of religious behaviour than each of the two measures by themselves. That is, this implicit measure of religiosity has incremental validity; it does not just serve as an alternative to its self-report counterpart, but can also supplement it.

As it turns out, partially-structured measures are not foreign to cognitive scientists of religion. Almost a decade before Vargas et al.’s (2004) paper, Barrett and Keil (1996) employed a similar technique to measure particular theological—in this case, anthropomorphic—beliefs. Adapting Bransford and McCarrall’s (1974) story comprehension paradigm, they presented participants with eight vignettes, such as the following:

A boy was swimming alone in a swift and rocky river. The boy got his left leg caught between two large, gray rocks and couldn’t get out. Branches of trees kept bumping into him as they hurried past. He thought he was going to drown and so he began to struggle and pray. Though God was answering another prayer in another part of the world when the boy started praying, before long God responded by pushing one of the rocks so the boy could get his leg out. The boy struggled to the river bank and fell over exhausted. (Barret and Keil 1996: 224)

Note that the vignette neither states nor necessarily implies that God is at one particular place at any given time, or that God moves at any point, or
even that God finished answering another prayer before saving the drowning boy. So, Barrett and Keil (1996) reasoned, to the extent that participants nevertheless later “remembered” that these pieces of information were present in the vignette, they were displaying anthropomorphic beliefs about God. In their series of experiments, Barrett and Keil (1996) found a disjunction between participants’ self-reported theological beliefs (e.g., in God’s atemporality, omnipresence) and their allegedly implicit theological beliefs that they used during these reading comprehension tasks.

So far, partially structured measures seem like promising methodological tools, adaptable for measuring many aspects of religious attitudes, from level of religious commitment to belief in supernatural agents to beliefs about supernatural agents. However, as for most of the techniques described in this paper, partially structured measures are still under-studied and under-utilized in contemporary research on religion.

2 Co-opting the Assimilation Bias

This next species of implicit measures have yet to be used in contemporary research, though as we’ll see, one recent experiment comes close. Saucier and Miller (2003) designed their Racial Argument Scale (RAS) based on the well-documented assimilation bias: the tendency to favourably interpret and evaluate information that supports one’s own beliefs. The RAS consists of a series of short paragraphs, presenting arguments with pro-Black conclusions and arguments with anti-Black conclusions. The participants’ task is to rate how well each argument supports its conclusion. Note that participants are not rating how much they agreed with the conclusions; in philosophical parlance, we might say that the participants are required to judge the arguments’ validity, rather than their soundness. In their series of studies, Saucier and Miller (2003) established that the RAS is internally consistent and reliable over time; furthermore, it predicts prejudiced behaviour and is moderately correlated with other self-report measures of racism while being less afflicted by the social desirability bias.

Now, while no one has adapted this paradigm for research on religion, Norenzayan and Hansen (2006: Experiment 2) come close. They were interested in the effects of thinking about death (i.e., mortality salience) on belief in supernatural agency. To measure belief in supernatural agency, Norenzayan and Hansen (2006: Experiment 2) presented participants with a New York Times article about an experiment on the efficacy of prayer, which found that women who were prayed for were far more likely to get pregnant than those who were not prayed for. After reading the article, participants were asked to rate their agreement to a series of statements. Some of these statements referred to supernatural agents (e.g., “God/a higher power exists,” “the experiment offers evidence that God/a higher power can answer prayers”); others did not refer to supernatural agents (e.g., “The study was scientifically rigourous”). Now, Norenzayan and Hansen (2006: Experiment 2)
Experiment 2) only used the former items in their measures of belief in supernatural agency. They did not analyze the results of the latter items. However, Saucier and Miller’s (2003) recent work suggests that those items might serve as good covert measures of belief in supernatural agents.

**IMPLICIT MEASURES, PART 2: RESPONSE LATENCIES**

So far, we’ve looked at relatively low-tech implicit measures. However, most of the recent research on implicit measures has focussed on response latency and physiological measures. In the rest of this chapter, I shall discuss a variety of response latency measures, and almost completely ignore psychophysiological and neuroimaging techniques. As useful as EMG, EEG, fMRI and their ilk may be as measures of emotion, and despite some recent research on the neural correlates of religious experience (see McNamara 2006 for a collection of reviews) the neurosciences of “belief” and more specifically of “religious belief” (and more specifically still, religious belief as a psychological state, rather than a stable trait) are still very underdeveloped. As such, neuroscience has little to offer us methodologically, at least insofar as we are looking for implicit measures of religious belief. Instead, we shall turn to a battery of response latency measures.

The basis of every response latency measure described here is the binomial choice reaction time (CRT) task. As its name implies, participants are presented with stimuli, which they have to categorize as quickly as possible. For example, the lexical decision task requires participants to decide whether a string of letters is a word or a non-word. Other examples we will encounter later include categorizing words as positive or negative (e.g., Fazio et al. 1995), and categorizing faces as George W. Bush or John Kerry (e.g., Nosek and Hansen 2008). The underlying assumption of these response latency paradigms is that differences in reaction time indicate differences in mental processes, such as concept accessibility, level of activation, or association between concepts. Rather than attempting to discuss the meaning of faster or slower reaction times abstractly and generally, however, we shall look at how they are used in specific tasks.

**1 Binomial Choice Reaction Time Test of Religious Belief**

The first response latency task simply consists of a binomial choice reaction time task, as described above. Cohen, Shariﬀ, and Hill (2008) presented participants with a series of nouns, which participants had to categorize as quickly as possible as either “real” or “imaginary.” Some of these nouns referred to things that were uncontroversially real (e.g., chair, George Bush), others referred to things that were uncontroversially imaginary (e.g., Bugs Bunny, Easter Bunny), others referred to objects of religious faith (e.g., God, miracles), and yet others referred to objects of “secular faith” (e.g., black...
hole, Abraham Lincoln). The items of interest were, of course, the objects of religious faith. Indeed, there were no statistically significant relationships between self-reported religiosity and choice reaction time for the other kinds of objects. For the objects of religious faith, however, Cohen et al. (2008) found a curvilinear relationship between self-reported religiosity and response latencies. That is, very religious people categorized objects like “God” as “real” faster than did more nominally religious people; and very non-religious (perhaps even anti-religious) people categorized objects like “God” as “imaginary” faster than did more nominally non-religious people. Cohen et al. (2008) interpreted these results as indicating that very religious and very non-religious people had highly accessible religious beliefs, whereas their more lukewarm counterparts has less accessible religious beliefs. However, a subtly different interpretation—in terms of association strength between concepts—is possible: Cohen et al.’s (2008) results might indicate that the concepts “real” and “God” are more closely associated in the minds of very religious people than in the minds of nominally religious people, and conversely that the concepts “imaginary” and “God” are more closely associated in the minds of militant atheists than in the minds of apathetic non-religious folk (cf. Gawronski and Bodenhausen 2007). In other words, response latencies might be indicative of the strength of one’s religious beliefs, not just the extent to which their religious beliefs are accessible or salient. If so, this CRT measure can obviously be used to study contextual influences on religious belief, factors that increase or decrease stated religious belief. In support of this interpretation, Jong, Halberstadt, and Bluemke (2012) recently found that an adapted version of this binomial choice reaction time task was sensitive to mortality salience priming.

Despite their simplicity, such minimalistic CRT paradigms have not caught on among experimental social cognitivists. This is potentially due to the interpretive ambiguity discussed above. So, most response latency paradigms are more complicated than just consisting of a single choice reaction time task. In the rest of the chapter, we will explore two genera of these paradigms, looking at widely-used examples of each, and considering and their potential applications in the cognitive science of religion.

2 Priming Measures

The first genus of commonly used response latency tasks exploits the effects of priming on participants’ reaction times. These tasks are structurally similar, in that participants are first presented with a prime, followed by a target, to which participants have to respond. In the most general terms, the effect of a prime on a participant’s response to a target is taken to indicate something about the associations between the corresponding concepts in the participant’s mind. But again, let’s look at some specific tasks.

Early in the piece, Meyer, Schvaneveldt, and Ruddy (1975) looked at the effects of semantic priming on participants’ response latencies in lexical
decision tasks. Whereas in most contemporary research, the primes are presented briefly and do not require responses, Meyer et al. (1975) simply had their participants make a series of word/non-word judgements. Ignoring the non-word trials, the target word would sometimes be semantically-related to the target word that immediately followed; at other times, the two target words would not be semantically related. As expected, participants responded faster when the preceding word was semantically related to the current target word.

More recent research has demonstrated the flexibility of this basic paradigm. For example, Fazio et al. (1986) replaced the lexical decision task with an adjective categorization task. They primed participants with attitude objects (e.g., hell) before presenting them with target adjectives (e.g., awful), which participants had to categorize as either positive or negative. As expected, participants’ responses were faster when the word pairs had similar connotations, than when they were incongruent.

More recently, Wittenbrink, Judd, and Park (1997) subliminally primed participants with the words “Black” or “White,” before each trial of a lexical decision task involving stereotype words (e.g., athletic, poor, intelligent, materialistic). They found that “Black” priming facilitated responses to negative stereotypes more than to positive ones, and vice versa for “White” priming. Furthermore, this effect was correlated with scores on a self-report measure of racial attitudes.

The flexibility of this sequential priming paradigm suggests that psychologists interested in religion may be able to adapt it to measure belief in and about supernatural agents. So far, although there have been a few published experiments looking at the effects of religious priming on various psychological phenomena like prosociality (Pichon, Boccato, and Saroglou 2007; Shariff and Norenzayan 2007), attribution of event authorship (Dijksterhuis et al. 2008), and humour creation (Saroglou and Jaspar 2001), there have not been any successful attempts at designing a priming measure of religious belief. Wenger’s (2004) work, however, is certainly a step in the right direction. Wenger (2003) had previously found that subliminally priming Christian participants with religious words (e.g., bible, christ, heaven, sermon) significantly increased the likelihood of participants reporting at least one biblical event when they were asked what they thought the three greatest events in the history of the world were. Wenger (2004) then primed participants with the words “Christian” (religious prime), “student” (neutral prime), and “housetop” (non-human neutral prime) and measured participants’ response latencies to categorize a series of actions as possible to perform or impossible to perform. This categorization task is somewhat analogous to the lexical decision task described above, but instead of words and non-words, the stimuli consisted of religious behaviours (e.g., worship God), academic behaviours (e.g., take tests), and non-behaviours (e.g., climb grass). In this study, Wenger (2004) found that the religious prime facilitated responses to the religious behaviours to a greater
extent for participants who self-reported high levels of intrinsic religiosity. However, as no correlational analyses were reported, it is unclear if this paradigm provides us a useful continuous measure of religiosity. Furthermore, given the relatively homogenous Christian sample in Wenger’s (2004) study, it remains to be seen whether this task will be a useful measure in a more heterogenous sample. Still, Wenger’s (2004) task is potentially useful as a measure of intrinsic religiosity, and may be adapted to measure other religious attitudes (e.g., belief in supernatural agents).

3 Dual Task Measures

The second genus of widely-used response latency tasks consists of dual task measures, which exploit interference effects between two tasks. Structurally, dual task measures—as the name suggests—involve two tasks, the performance in at least one of which is affected by some feature of the other. The prototypical example of a dual task measure in social cognitive research is the Implicit Association Test (IAT; Greenwald, McGhee, and Schwartz 1998). However, the Stroop Task (Stroop 1935) serves as a good introduction to dual task measures.

The Stroop Task requires participants to name the colour in which a stimulus is presented. The stimulus may be a shape, or a random string of letters, or a word, or the name of the colour it’s presented in, or the name of another colour. What Stroop (1935) showed, was that people find it very difficult to name the colour of a stimulus, when the stimulus is another colour word. For example, we find it very difficult to name the colour of the word “brown” presented in “blue.” Not only do we make more errors, but we are also slower when we succeed, than with other kinds of stimuli. What seems to be going on is that an automatic process—reading—is interfering with the task: naming the colour.

Research on the Stroop effect has continued unabated since 1935; over the last seven decades, it has become clear that the Stroop Task is also rather flexible. The most well-known expansion of the Stroop Task regards the emotional Stroop effect: for example, anxious people take longer to name colours of negatively-valenced, emotionally-charged words (Williams, Mathews, and MacLeod 1996); and patients with eating disorders take longer to name colours of food-related words (Ben-Tovim et al. 2006). More recently, Nicholas Gibson (2005) attempted to design a religious Stroop Task. He found that evangelical Christians experienced significantly more interference for religious words than did atheists, but the effects were not large enough to serve as a useful measure of religiosity.

Research on adapting the IAT for our purposes have been more promising. The classic IAT basically consists of two different categorization tasks put together in five different configurations. In Phase 1, participants are presented with a series of targets (e.g., images of George Bush and John Kerry; Nosek and Hansen 2008), which they have to categorize via key-presses. In
Phase 2, they are presented with a series of attribute words (e.g., positive and negative words) which they have to categorize using the same keys. In Phase 3, participants perform both categorization tasks simultaneously. That is, target categorization and attribute categorization trials are interspersed with each other, and the same keys are used for both trial types. In Phase 4, the target are presented again, but this time the categorization keys are reversed. Phase 5 is another mixed phase, differing from Phase 3 only in the reversal of the target categorization keys. In keeping with our example, note that in Phase 3, “George Bush” shares a key with positive words, while in Phase 5 “George Bush” shares a key with negative words. As you might expect, the difference in average reaction times for Phase 3 and Phase 5 are indicative of the participant’s evaluation of George Bush. For example, an avid George Bush supporter (let’s call them Bushists) might perform very quickly in Phase 3 and very slowly in Phase 5, a nominal Bushist might not show much of a difference between phases, and a radical anti-Bushist might respond very quickly in Phase 5 and very slowly in Phase 3.

Thus far, there have been three published adaptations of the IAT for research on religion. Wenger and Yarbrough’s (2005) study looked at the relationship between self-reported intrinsic/extrinsic religiosity and implicit associations between self/other and intrinsic/extrinsic religiosity, and found that the explicit and IAT measures of intrinsic/extrinsic religiosity were significantly correlated. However, as the study was not specifically about religious beliefs, I shall not discuss it further.

Bassett et al. (2005) designed two IAT tasks, with religious words (e.g., god, bible) or spiritual words (e.g., meditation, tranquility) as targets respectively, and positive/negative words as attribute words. Rather than a standard IAT, it appears that they used a modified version of the IAT called the Single-Target Implicit Association Test (ST-IAT; Wigboldus, Holland, and van Knippenberg 2006). In contrast to the standard IAT, the ST-IAT consists of three phases. Phase 1 is an attribute categorization task; Phase 2 couples the attribute categorization task with a simple reaction time task, in which they press a key (shared with one of the attribute categories) whenever a target is presenting; Phase 3 is identical to Phase 2, except that the key for the simple reaction time task is switched, to be mapped on to the other attribute category. Although research on and utilizing the ST-IAT is still relatively sparse, Bluemke and Friese (2008) found that, if cautiously used, the ST-IAT can be a psychometrically reliable and valid measure of attitudes toward the target.

Bassett et al. (2005) also recorded self-report data on religiosity and spirituality, generating five religiosity and spirituality scores. They found that scores for the two IATs were significantly correlated; and that scores from several of the different self-report measures were significantly correlated with one another; but that only a single-item Christian identity measure was significantly correlated with the religious IAT. Despite Bassett et al.’s (2005) results, I don’t think we should be discouraged about our prospects.
for adapting the IAT for research on religion. The religious IAT has not yet been tested on a more religiously heterogeneous sample; judging by the results of their self-report measures, Basset et al.’s (2005) participants were all highly religious Christians. Perhaps the religious IAT is not sensitive enough to measure slight variations in religiosity, but it doesn’t have to be to be a useful measurement tool. Furthermore, it is unclear how Bassett et al. (2005) selected the target stimuli. On the basis of their experimental work, Bluemke and Friese (2006) rightly caution IAT users of the potentially confounding effects of features of the stimuli we use. In Bassett et al.’s (2005) study, many of the items in their spiritual IAT (e.g., relationship, tranquility, meditation, inner peace) seem to carry strong positive connotations independently of their semantic connection to spirituality. That is, even militant atheists will evaluate such items positively, regardless of their attitudes toward spirituality. Even the religious IAT might suffer from this problem, albeit to a lesser extent. If so—and of course, this is an empirical question—Bassett et al.’s (2005) IATs are of limited value as measures of attitudes toward religion and spirituality.

More recently, a somewhat more sophisticated and successful attempt to use an ST-IAT as an implicit measure of religiosity was executed by Shariff, Cohen, and Norenzayan (2008). For their Implicit Religiosity IAT (IR-IAT), Shariff and Norenzayan (2008) used religious words (e.g., god, devil) as targets, and synonyms for true (e.g., true, real) and false (e.g., false, bogus) as attributes; and found that IR-IAT scores significantly correlated with self-report religiosity measures. Furthermore, IR-IAT scores were sensitive to manipulation: Participants who were first exposed to a paragraph arguing against the existence of God led to decreased religiosity, on both the IR-IAT and the self-report measure.

GO FORTH AND DO LIKEWISE

With the exception of Barrett and Keil’s (1996) theological anthropomorphism paper, all the applications of implicit measures to the cognitive science and psychology of religion I reviewed today were published in and after 2005. Relative to psychologists who specialize in the study of prejudice or self-esteem, we are neonates to the world of implicit measures, and measures of implicit emotion and cognition. So far, we have a handful of experiments that demonstrate the great potential and utility of such measures for the much-needed empirical work in our area. But, of course, we have a long way to go. First, we need to get comfortable with these techniques and use them more. In the next edition of Hood, Hill, and Spilka’s (2009) widely used The Psychology of Religion: An Empirical Approach, I’d like to see more than one page (of its 487) devoted to implicit measures. After all, Pete Hill is involved in this research; indeed, the CRT measure described earlier (Cohen et al. 2008) emerged out of
his earlier, unpublished work (Hill, Jennings, and Haas 1992). Second, we need to join in the effort to understand these implicit measures better. There are many thorny methodological and conceptual issues, which I have hinted at, but not fully discussed in this paper. Psychologists and cognitive scientists of religion are, as I submitted earlier, interested in people’s beliefs in and about supernatural agents. But how does a CRT score or an IAT score relate to this mental state—belief—which we have perhaps naively stolen from folk psychology? How do we make sense of implicit beliefs or, more provocatively, unconscious beliefs? How do these relate to the concept of intuitive beliefs, which most of us have accepted into our vocabularies (Sperber 1997). Indeed, what is a belief? These are questions that psychologists have been wrestling with for decades, in different contexts. And now, we—those of us who are committed to the study of religion—have to join in this discussion.

REFERENCES


Jong, J., M. Bluemke, and J. Halberstadt. Submitted. “Fear of Death and Supernatural Beliefs: Developing a New Supernatural Belief Scale to Test the Relationship.”


