

# Oracle of the unconscious or deceiver of the unwitting?

Aiden P. Gregg on the implicit association test

**Over the last 10 years, the implicit association test (IAT) has held considerable appeal to applied researchers as an indirect index of what people are thinking and feeling. It dependably yields large effects, but debate has raged over what these effects actually mean. Does the test cut through the fog of consciousness, or is it little more than a cognitive curiosity?**

**This article outlines what might be behind both dissociations and associations between implicit and explicitly measured attitudes. It asks which represent our 'real' attitudes, and whether your IAT results really matter.**

Suppose you are an educated Western liberal, now non-religious despite a traditional Judaeo-Christian upbringing. You pride yourself on being unprejudiced. In particular, you regard it as both unjust and irrational to discriminate against, or even think ill of, individuals just because they are members of a social group, especially one traditionally held in low esteem.

Now imagine that as part of a psychology study, you are invited to complete a special type of computerised classification task. In this task, you must assign items to one of four categories by pressing one of two keys. You must also do this as quickly as you can without making errors. The four categories in question are Christians, Muslims, Good and Bad. Each category has a corresponding set of items: Jesus, Trinity, Bible; Muhammad, Allah, Koran; nice, kind, excellent; and nasty, mean, awful.

You correctly guess that the purpose of the task is to assess, in some newfangled way, your attitudes towards Christians and Muslims. To get ready, you remind yourself what those attitudes are.

Certainly, Islam has received much bad press lately. Most extremists launching suicide attacks have been Muslims. However, you note that most Muslims neither engage in, nor politically support, such attacks, making it unfair to blame the many for the misdeeds of the few. You also

recognise that Christianity, for all its modern moderation, has had its share of past outrages. Finally, you recognise that, despite some illiberal teachings about women and homosexuality, both Christianity and Islam do strive to uphold important values, such as charity and solidarity. You decide that Muslims and Christians are, on the whole, moderately good, with neither being better or worse overall than the other.

And so the task begins. You have been informed that it will consist of two blocks (see Figure 1 for a schematic outline). In the first block, the categories Christians and Good appear on the upper left of the screen, and the categories Muslims and Bad on the upper right. The items appear below, one after the other, in the middle of the screen. Items belonging to the former

pair of categories are to be classified by pressing a key on the left ('Q'), and those belonging to the latter pair by pressing a key on the right ('P').

After a few tentative keystrokes, and the

occasional error, you get into your stride. You had expected it would be awkward to classify four types of item using only two keys. However, it turns out to be straightforward. You complete the first block without even breaking a sweat.

Then the second block begins. You feel confident that this will pose no problems either. True, the names of the faiths have switched sides: Christians and Bad are now on the left, and Muslims and Good on the right. Still, you reckon that this should make no difference: your feelings towards Muslims and Christians are indistinguishable, and the task merely requires you to reclassify the same sets of words.

However, that is not how things pan out. Your key pressing, so fluid and flawless in the previous block, has become hesitant and halting. You find yourself, again and again, pressing the wrong key by accident. In the end, you are glad when the

"...it is one of the few objective psychological tests to attract attention outside academia"

## questions

Can a reaction time task reveal facts about people that they don't realise are true, don't want to share with others, or don't want to admit to themselves?

## resources

[http://faculty.washington.edu/agg/iat\\_validity.htm](http://faculty.washington.edu/agg/iat_validity.htm)  
Nosek, B.A., Greenwald, A.G. & Banaji, M.R. (2007). The Implicit Association Test at age 7: A methodological and conceptual review. In J.A. Bargh (Ed.) *Automatic processes in social thinking and behavior* (pp.265–292). New York: Psychology Press.

## references

- Arkes, H.R. & Tetlock, P.E. (2004). Attributions of implicit prejudice, or "Would Jesse Jackson fail the Implicit Association Test?" *Psychological Inquiry*, 15, 257–278.
- Blanton, H. & Jaccard, J. (2006). Arbitrary metrics in psychology. *American Psychologist*, 61, 27–41.
- Chaiken, S. & Trope, Y. (1999). *Dual process theories in social psychology*. New York: Guilford Press.
- Cunningham, W.A., Preacher, K.J. & Banaji, M.R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science*, 12, 163–170.
- De Houwer, J., Geldof, T. & De Bruycker, E. (2005). The implicit association test as a general measure of similarity. *Canadian Journal of Experimental Psychology*, 59, 228–239.
- Fiedler, K. & Bluemke, M. (2005). Faking the IAT. *Basic and Applied Social Psychology*, 27, 307–316.
- Fiedler, K., Messner, C. & Bluemke, M. (2006). Unresolved problems with the "I", the "A", and the "T". *European Review of Social Psychology*, 17, 74–147.
- Gawronski, B. & Bodenhausen, G.V. (2006). Associative and propositional processes in evaluation. *Psychological Bulletin*, 132, 692–731.
- Gawronski, B., LeBel, E.P. & Peters, K.R. (2007). What do implicit measures tell us? *Perspectives on Psychological Science*, 2, 181–193.
- Greenwald, A.G., McGhee, D.E. & Schwartz, J.K.L. (1998). Measuring individual differences in implicit cognition. *Journal of Personality and Social Psychology*, 74, 1464–1480.
- Greenwald, A.G., Nosek, B.A. & Banaji, M.R. (2003). Understanding and

block concludes. It was quite a chore.

You have just completed an implicit association test, or IAT (Greenwald et al., 1998; Nosek et al., 2007). The brainchild of Dr Anthony Greenwald from the University of Washington, it is one of the few objective psychological tests (that is, not involving self-report) to attract attention outside of academia. So what exactly is the IAT all about?

### What the IAT is designed to measure

As its name suggests, the IAT is designed to test for the presence of implicit associations. The adjective 'implicit' here means two things: first, that these associations elude conscious introspection; and second, that they defy conscious control. That is, implicit associations are both unconscious and automatic.

In the scenario described, you took longer to accurately complete the second block than the first. This indicates that, somewhere in your mind, Christians–Muslims maps better on to Good–Bad than on to Bad–Good, or more simply put, that

you exhibit an implicit preference for Christians over Muslims. Moreover, the fact that, despite your best efforts, you couldn't avoid going slower in the second 'incompatible' block than the first 'compatible' one suggests that the associations operated automatically. In addition, the fact that the effect took you by surprise suggests that these associations existed unconsciously.

The scenario above is illustrative, but many real IATs have yielded comparable results. For example, over the last few years, thousands of people have taken an online version of a race IAT at Project Implicit (<https://implicit.harvard.edu/implicit/>). This IAT features the categories Good, Bad, White and Black. Respondents classify positive or negative words into the first pair, and lighter or darker faces into the second. Although few white respondents report preferring their ingroup, roughly 80 per cent of them show an implicit preference for White on this IAT, often to their great chagrin (Nosek et al., 2002). Black respondents, in contrast, emerge as more implicitly egalitarian: only about 50 per cent show a corresponding directional preference for Black.

By substituting different categories into an IAT, a variety of psychological variables can be assessed implicitly (see Lane et al., 2007). These include sexism (e.g. Man; Woman; Rational; Emotional), homophobia (e.g. Straight; Gay; Virtuous; Wicked), self-esteem (e.g. Me; Not-Me; Good; Bad), gender identity (e.g. Me; Not-Me; Male; Female), personality traits

(e.g. Me; Not-Me; Extraverted; Shy), and consumer attitudes (e.g. Coke; Pepsi; Tasty; Tasteless).

### The popularity of the IAT

Since its invention in the late 1990s, the IAT has been eagerly adopted by many applied researchers. This is partly because it offers them a way of bypassing the pitfalls of self-report. To collect data, applied researchers still mostly have to ask people questions: no other method of inquiry is as flexible or efficient. However, the answers that people provide can be misleading: people can lie outright; present themselves in a socially desirable light; deceive themselves that they hold the 'right' opinions; or provide haphazard responses on the spur of the moment. Moreover, the more sensitive the topic under investigation, the greater the risk of a biased response. Hence, the IAT holds considerable appeal as an indirect index of what people are thinking and feeling. It promises to circumvent conventional sources of error.

Note that the IAT is not alone: it is flanked by an ever-expanding array of implicit measures, each with its own distinctive modus operandi (Wittenbrink & Schwarz, 2007). However, unlike many other implicit measures, the IAT has a key virtue that ensures its continuing popularity: it dependably yields large effects. Indeed, as our hypothetical scenario illustrated, these effects are often large enough for respondents themselves to notice.

### Ideologies of the IAT

Although no one disputes that the IAT can be relied upon to yield large effects, controversy rages over what those effects mean. For the purposes of exposition, let us define the boundaries of the debate by caricaturing the positions taken by opposing partisans. In reality, most researchers take a more nuanced view.

One position is IAT Adoration. The IAT is hailed, not only as a remedy for self-



Figure 1: Schematic illustration of the two blocks of a Christians–Muslims IAT.

using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, 85, 197–216.

Greenwald, A.G., Nosek, B.A. & Sriram, N. (2006). Consequential validity of the implicit association test: Comment on the article by Blanton and Jaccard. *American Psychologist*, 61, 56–61.

Greenwald, A.G., Poehlman, T.A.,

Uhlmann, E. & Banaji, M.R. (in press). Understanding and using the implicit association test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*.

Gregg, A.P., Seibt, B. & Banaji, M.R. (2006). Easier done than undone: Asymmetry in the malleability of implicit preferences. *Journal of Personality and Social Psychology*, 90, 1–20.

Hofmann, W., Gawronski, B., Gschwendner, T. et al. (2005). A meta-analysis on the correlation between the implicit association test and explicit self-report measures. *Personality and Social Psychology Bulletin*, 31, 1369–1385.

Karpinski, A. & Steinman, R.B. (2006). The single category implicit association test as a measure of implicit social cognition. *Journal of*

*Personality and Social Psychology*, 91, 16–32.

Kim, D.Y. (2003). Voluntary controllability of the implicit association test (IAT). *Social Psychology Quarterly*, 66, 83–96.

Lane, K.A., Banaji, M.R., Nosek, B.A. & Greenwald, A.G. (2007).

Understanding and using the implicit association test: IV. What we know (so far). In B. Wittenbrink & N.S. Schwarz (Eds.) *Implicit measures of*

report bias, but also as a means of moral enlightenment. In effect, IAT cuts through the fog of consciousness to what people really think and feel. Unfortunately, that reality is often unpalatable: apparently open-minded people can be bigots beneath. Such insidious bigotry must be widely publicised if society is ever to rid itself of the scourge of prejudice.

The other position is IAT Antipathy. IAT effects are derided as mere cognitive curiosities, devoid of practical implications. The IAT is an interesting parlour game, but it reveals nothing of substance about a person's psychological make-up. All that matters is what people consciously believe and deliberately do. The real danger, rather, is that well-meaning people, upon receiving IAT feedback, will conclude that they harbour non-existent prejudice, and suffer misplaced guilt.

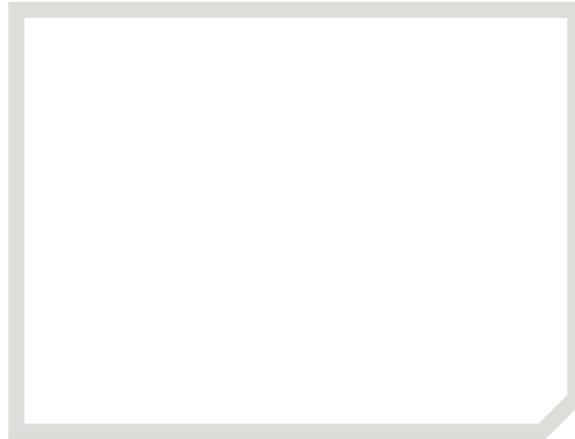
So, is the IAT the royal road to the unconscious or the primrose path to self-delusion?

### Explicit/implicit dissociations

A convenient way to enter the debate is to ask why explicit opinions (reflected in self-reports) and implicit associations (reflected in IAT scores) are often dissociated (Nosek, 2007). That is, why is the correlation between them typically so modest (mean  $r = .19$ : Hofmann et al., 2005)? For example, suppose I measure your self-esteem using both a standard questionnaire and a custom-built IAT. Although both indices will register a positivity bias (people generally like themselves, both explicitly and implicitly), knowing one index will not really assist me in predicting the other. Why not?

One explanation is that explicit opinions and implicit associations derive from separate sources. Abundant theory, and some evidence, point to the human mind being divided into two largely independent subsystems: first, a familiar foreground, where processing is conscious, controlled, intentional, reflective and slow,

but where learning occurs rapidly; and second, a hidden background, where processing is unconscious, automatic, unintended, impulsive and fast, but where learning occurs gradually (Chaiken & Trope, 1999; Searle, 1992; Strack, &



What if your explicit opinions and automatic associations are at odds?

Deutsch, 2004). Now, if explicit opinions are foreground-based, and implicit associations are background-based, then they too should be statistically independent.

This dual-process model is appealingly neat. It recalls Freud's model of the mind, but with the inner sex maniac replaced by a dull but efficient zombie. However, there more mundane explanations for the existence of explicit/implicit dissociations.

Measurement unreliability is one possibility. The IAT only gives moderately consistent results (mean  $\alpha = .79$ : Hofmann et al., 2005). Simply put, if you do the same IAT several times, you get somewhat different results each time (although alternative implicit measures fare worse). The upshot is that the 'real' size of correlations between IAT scores and questionnaire scores may be underestimated.

However, measurement unreliability

cannot be the whole story. Even when its deflationary impact is statistically removed, a substantial explicit/implicit dissociation remains (Cunningham et al., 2001). Hence, the IAT does measure something distinguishably different from self-report measures, whatever that is.

Another possibility is that explicit/implicit dissociations reflect a lack of conceptual correspondence between what IATs and self-report measures assess (Hofmann et al., 2005). Consider again the social categories Christians and Muslims. Both are very broad and comprise several subtypes. For example, Christians could mean easygoing Anglicans or born-again Baptists, and Muslims could mean radical Shias or mystical Sufis. Clearly, different subtypes will come to mind, depending on which IAT stimuli and self-report items are used and on how respondents interpret them. But unless an IAT and self-report measure

both evoke the same target, no strong correlation would be expected.

The IAT also has a procedural limitation that can threaten conceptual correspondence: it assesses implicit associations towards one category relative to another. But what if a given category has no natural contrast, or has several possible contrasts? For example, does the 'Me' category in the self-esteem IAT really require a contrast? And if one is used, should it be 'Not-Me', 'Other', 'Them', 'Friend', or 'Enemy'? One might argue that all judgements are comparative, so that having to specify a contrast on the IAT is not a genuine limitation. Nonetheless, if an IAT and self-report measure do not specify the same contrast, then their correlation will diminish. To get around the specificity problem, some researchers have designed alternative IATs in which the contrast category is dropped or de-emphasized (e.g. Karpinski, & Steinman, 2006; Nosek &

*attitudes* (pp.59–102). New York: Guilford Press.

Lowery, B., Hardin, C. & Sinclair, S. (2001). Social influence effects on automatic racial prejudice. *Journal of Personality and Social Psychology*, 81, 842–855.

Maison, D., Greenwald, A.G. & Bruin, R. (2001). The implicit association test as a measure of implicit consumer attitudes. *Polish Psychological*

*Bulletin*, 2, 61–79.

Maison, D., Greenwald, A.G. & Bruin, R.H. (2004). Predictive validity of the implicit association test in studies of brands, consumer attitudes, and behavior. *Journal of Consumer Psychology*, 14, 405–415.

Nosek, B.A. (2007). Implicit-explicit relations. *Current Directions in Psychological Science*, 16, 65–69.

Nosek, B.A. & Banaji, M.R. (2001). The

go/no-go association task. *Social Cognition*, 19, 625–666.

Nosek, B.A., Banaji, M.R. & Greenwald, A.G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration website. *Group Dynamics*, 6, 101–115.

Nosek, B.A., Greenwald, A.G. & Banaji, M.R. (2005). Understanding and using the implicit association test: II. Method variables and construct

validity. *Personality and Social Psychology Bulletin*, 31, 166–180.

Nosek, B.A., Greenwald, A.G. & Banaji, M.R. (2007). The implicit association test at age 7. In J.A. Bargh (Ed.) *Automatic processes in social thinking and behavior* (pp.265–292). Hove: Psychology Press.

Nosek, B.A. & Hansen, J.J. (2008). The associations in our heads belong to us. *Cognition and Emotion*, 22,

Banaji, 2001). The price they pay, however, is somewhat smaller effects.

### Explicit/implicit associations

Yet sometimes the IAT does correlate well with self-report measures (Nosek, 2007). This often happens, for example, for political candidates (e.g. Bush vs. Gore: Nosek et al., 2002) and consumer brands (e.g. Apple vs. Microsoft: Maison et al., 2001). Such high correlations suggest one or other of the following. First, the background and foreground of the mind can be in alignment, even if they tend not to be. Second, both the IAT and self-report measures can tap into either the background or foreground, even if either measure mostly taps into one or the other. Put more simply, the IAT can sometimes pick up what is currently on your mind, and self-report measures what currently isn't. Hence, what the IAT reflects need not always be implicit in the sense of being unconscious (Gawronski et al., 2007). The dual-process account is therefore a little rough around the edges.

More practically, several factors have been found to predict stronger correlations between the IAT and self-report measures (Hofmann et al., 2005; Nosek et al., 2005). One, perhaps, deserves to be singled out: attitude strength. If you feel strongly about something, then your questionnaire responses and IAT scores are more likely to harmonise. Both the background and the foreground of your mind are working in concert.

### The reality question

But what if your explicit opinions and automatic associations are at odds? To return to our earlier scenario, suppose you consciously believe that Christians and Muslims are equally good, but show a strong implicit preference for Christians over Muslims. What, then, is your real attitude?

Someone disposed to IAT Adoration might say that your implicit associations represent the reality. After all, self-reports

are known to be biased in various ways. In contrast, automatic and unconscious attitudes cannot, by definition, be faked. More philosophically (and here orthodox Freudians would agree), the profound, mysterious and hidden is always more telling than the superficial, commonplace and obvious. So, contrary to what you consciously believe, you actually like Christians more than you like Muslims.

In contrast, someone disposed to IAT Antipathy might say that only your explicit opinions matter. After all, for an attitude to be real, it must first be endorsed. But such endorsement requires truth to be distinguished from falsity, something that can only occur in the mental foreground (Gawronski et al., 2007; Strack & Deutsch, 2004). Hence, whatever implicit associations may lurk in the background, it is neither here nor there. Indeed, implicit associations may just reflect your mind taking note of attitudes that are socially prevalent. For example, most Westerners recognise that negative attitudes towards Muslims exist, even if they personally disagree with them. Perhaps IAT scores merely highlight such passive recognition of what others think and feel (Arkes & Tetlock, 2004; but see Nosek & Hansen, 2008).

A compromise view is that personal attitudes can be jointly composed of explicit opinions and automatic associations. But if so, what might the latter tell us that the former cannot?

### Antecedents and consequences of IAT effects

For IAT effects to mean anything, they must relate to something beyond themselves. In particular, they must be preceded (or caused) by meaningful antecedents and followed by (or cause) meaningful consequences. 'Meaningful' here means either theoretically expected or practically significant. From a theoretical point of view, the dual-process

account implies that background processes should operate quickly but change slowly. If so, then, relative to self-report measures, the IAT might be expected to better predict spontaneous behaviours, and to yield effects that are relatively invariant in the short-term. Also, from a practical point of view, the IAT should best predict behaviour in sensitive domains where self-reports are likely to be biased. What does the research say?

Reassuringly, the IAT does predict behaviour: a meta-analysis of over 100 relevant studies confirms it (Greenwald et al., in press; Lane et al., 2007). This proves that IAT effects are not trivial. Admittedly, the level of prediction is modest.

Nonetheless, the behaviours documented are often quite specific, so it is striking that general implicit associations predict them at all. Moreover, the IAT outstrips self-report in forecasting instances of discrimination and prejudice. Hence, it offers some genuine diagnostic advantages.

The pattern of prediction shown, however, is messy. Sometimes – in keeping with the dual-process account – the IAT predicts spontaneous behaviours better while self-report predicts deliberate behaviours better. Other times, however, the IAT and self-report predict the same outcome – either independently, redundantly, or interactively (Maison et al., 2004; Perugini, 2005). Recent evidence suggests that the unique predictiveness of the IAT can be maximised by simultaneously taking into account various situational triggers of spontaneous behaviour (Perugini & Prestwich, 2007).

Switching to antecedents, many factors, and often surprisingly subtle ones, have been found to influence IAT effects. For example, the standard IAT preference among whites for White over Black gets weaker in the presence of a black experimenter (Lowery et al., 2001). Perhaps respondents, seeking to appear politically correct, intentionally adjust their performance. Alternatively, spending time with an individual black person may dilute a respondent's negative stereotypes.

So, can respondents really fake IAT performance? And can very recent experiences shape IAT effects? In both cases, the answer appears to be a qualified yes. As regards faking, IAT effects persist when naive respondents are warned about it beforehand (Kim, 2003). However, respondents who understand how the IAT works can learn to successfully conceal their implicit attitudes (Fiedler &

“automatic and unconscious attitudes cannot, by definition, be faked”

553–594.

Perugini, M. (2005). Predictive models of implicit and explicit attitudes. *British Journal of Social Psychology*, 44, 29–45.

Perugini, M. & Prestwich, A. (2007). The gatekeeper. *European Journal of Personality*, 3, 303–317.

Rothermund, K. & Wentura, D. (2004). Underlying processes in the implicit association test (IAT). *Journal of Experimental Psychology: General*, 133,

139–165.

Searle, J. (1992). *The rediscovery of the mind*. Cambridge, MA: MIT Press.

Strack, F. & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, 8, 220–247.

Wittenbrink, B. & Schwarz, N. (Eds.) (2007). *Implicit measures of attitudes*. New York: Guilford Press.

Bluemke, 2005). They achieve this mainly by deliberately slowing down in the 'compatible' block; deliberately speeding up in the 'incompatible' block is more difficult. Hence, although the IAT may not often be beaten in practice, it can be partly beaten in principle.

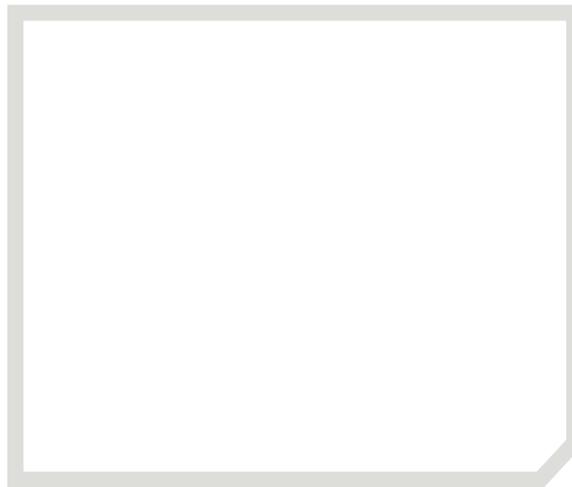
As regards recent experience, it is now known that implicit associations can form rapidly, sometimes on the thinnest of pretexts. For example, telling people a story about a nice group and nasty group suffices to create an implicit preference for the former over the latter – as does merely asking them to suppose that such groups exist (Gregg et al., 2006)! On the other hand, once formed, such implicit associations show more inertia than explicit opinions. Overall, the picture is confusing. Primitive types of learning (e.g. conditioning) do tend to shape implicit associations, and more sophisticated types (e.g. cognitive dissonance) may tend to shape explicit opinions; but it sometimes works the other way around too (Gawronski & Bodenhausen, 2006).

Again, the dual-process account – which implies that the IAT should assess slow-to-change dispositions that are responsible for spontaneous behaviour – is only approximately supported. Nonetheless, the IAT clearly does relate to some meaningful antecedents and consequences (Greenwald et al., 2006). So the question is not whether IAT effects matter, but how much they matter. One way of assessing how much is to inquire into the significance of individual IAT results.

### The IAT and you

Returning to our opening scenario, suppose you went 300ms faster in one IAT block (Christians & Good; Muslims & Bad) than in another (Christians & Bad; Muslims & Good). What would this result tell you about yourself? The answer, perhaps somewhat disappointingly, is not much for certain. Such a result is more suggestive than indicative. There are three reasons for this.

First, an individual IAT score cannot be interpreted in isolation: it requires some context. The most obvious one is other IAT scores. Generally speaking, 300ms would be considered a reasonably large score. However, some critics of the IAT



Telling people a story about a nice group and nasty group suffices to create an implicit preference for the former

point out that, unlike IQ tests, the IAT has not been formally standardised, nor are systematic norms available for quantifying relative performance (Fiedler et al., 2006). This is indeed a drawback, although one that could easily be remedied for any particular purpose.

Second, even if general links between the IAT and meaningful phenomena are demonstrated, they cannot be extrapolated to specific individuals. For example, if, in a group of individuals, an implicit preference for Christians over Muslims correlates with having more Christian friends, it does not follow that every individual with such a preference will have more Christian friends. Two factors ensure this. One (pointed out earlier) is measurement unreliability, which causes scores to fluctuate randomly. The other is the presence of systematic confounds – factors other than implicit associations that contribute to IAT effects. For example, older people show larger IAT effects than younger people do, perhaps because the cognitive skills required to complete the more difficult incompatible block decline over time. To reduce impact of individual differences, special data-reduction algorithms have been developed (Greenwald et al., 2003).

Nonetheless, IAT effects can sometimes be spuriously generated. For example, for most people doing an IAT, Young–Old maps better on to Good–Bad than on to Bad–Good. This suggests that youth is implicitly 'good' while age is implicitly 'bad'. However, Young–Old also maps on better to Word–Nonword than to Nonword–Word. Does this imply that youth is implicitly 'wordy' while age is implicitly 'nonwordy'? Surely not! The

explanation seems to be that the categories Bad and Nonword stick out more than the categories Good and Word do, making it more natural to classify them using the same key (Rothermund & Wentura, 2004). Researchers disagree about how pervasive such salience asymmetry effects are, and whether they explain away IAT effects. However, everyone agrees that IAT scores are not pure measures of implicit associations (De Houwer et al., 2005).

Third, the precise relation between observed IAT scores and underlying automatic associations remains unknown: in formal terms, the IAT's metric, like many in psychology, is arbitrary (Blanton & Jaccard, 2006). Suppose an IAT measures, albeit imperfectly, implicit racial prejudice. Then, all else equal, a higher scorer will be more likely to be implicitly prejudiced than a lower scorer. However, exactly how much this likelihood increases as IAT scores increase is unclear. In terms of speed, an IAT differential of 600ms may be twice as large as one of 300ms, but the strength of implicit associations (and of what they predict) need not vary in a similar proportion. Hence, feedback about IAT effects themselves (small; medium; large) should not be confused with feedback about the implicit associations they tend to reflect (weak; moderate; strong): the former can be precisely described, but the latter must be tentatively inferred.

### The middle ground

Both the caricatured positions we described – IAT Adoration and IAT Antipathy – fail to do justice to the empirical complexities. The IAT is not an infallible index of what people really think and feel deep down. Rather, the technique has both conceptual and methodological pitfalls. Nonetheless, the IAT is more than just a source of misleading or empty information. It predicts real behavioural outcomes, and at the very least serves an informative adjunct to self-report. Time and further testing will reveal its ultimate significance.



**Aiden P. Gregg** is at the University of Southampton  
A.Gregg@soton.ac.uk

