PERSPECTIVE

WIREs COGNITIVE SCIENCE WILEY

# Whether implicit attitudes exist is one question, and whether we can measure individual differences effectively is another

## Chandra Sripada

University of Michigan, Ann Arbor, MI, USA

**Correspondence**
Chandra Sripada, University of Michigan, Ann Arbor, MI 48109, USA.
Email: sripada@umich.edu

**Edited by:** Wayne Wu, Editor

**Abstract**

Questions about measurement of individual differences in implicit attitudes, which have been the focus so far in this exchange, should be distinguished from more general questions about whether implicit attitudes exist and operate in our minds. Theorists frequently move too quickly from pessimistic results regarding the first set of questions to pessimistic conclusions about the second. That is, they take evidence that indirect measures such as the implicit association test (IAT) disappoint as individual difference measures and use it to (mistakenly) suggest that people do not in fact have implicit attitudes directed at stigmatized groups. In this commentary, I dissect this mistake in detail, drawing key lessons from a parallel debate that has unfolded in cognitive science about "conflict tasks" such as the Stroop task. I argue that the evidence overall supports a nuanced conclusion: Indirect measures such as the IAT measure individual differences in implicit attitudes poorly, but they—via distinct lines of evidence—still support the view that implicit attitudes exist.

This article is categorized under:
    Psychology > Theory and Methods

**KEYWORDS**
conflict tasks, implicit attitudes, measurement, ontology, test-retest reliability

## 1 | INTRODUCTION

The exchange about indirect measures of implicit attitudes between Brownstein, Madva, and Gawronski (Brownstein et al., 2019; Gawronski et al., 2022) and Machery (Machery, 2022a, 2022b) has been spirited, informative, and productive. The focus of their debate is on *measurement of individual differences* in implicit attitudes. Machery says indirect measures such as the implicit association test (IAT) do not measure individual differences well, marshaling evidence that these tasks have poor test–retest reliability, poor predictive validity, and a lack of cohesion across tasks, among other problems. Brownstein, Madva, and Gawronski say that measurement properties of indirect measures are good enough when judged by the appropriate benchmarks or that they can be made better, especially by accounting for

subtle contextual factors. I generally do not strongly disagree with what these authors say, or at least I do not plan to disagree very much here.

The main issue I want to examine in this commentary concerns the relationship between questions about measurement of individual differences in a construct and more general questions about whether that construct exists and operates in our minds. There is a widespread misconception that these two sets of questions are necessarily tightly connected—that poor performance with respect to the first set of questions justifies pessimistic conclusions with respect to the second set ("If you cannot measure it, then it does not exist"). Thus, it is common to see commentators make a quick transition from the claim that indirect measures such as the IAT fail as individual difference measures to the claim, or at least the suggestion, that people do not in fact have implicit attitudes directed at stigmatized groups (I discuss examples in §3).[1]

In what follows, I explain in detail why this transition is unjustified. I argue that it is perfectly well possible in some cases to reach a "divergent vedict" in which, though measurement of individual differences in a construct is extremely poor, there is nonetheless sufficient overall evidence that the construct exists and operates in our minds. This mixed verdict thus represents a "conviction" on the narrow question of measurement of individual differences but a "vindication" on the broader question of ontology.

While a divergent verdict is a logical possibility, why take it seriously? Because this position is exactly where the field has landed in a parallel debate about so-called "conflict tasks" such as the Stroop task and Simon task. Interestingly, the IAT and several other indirect measures *are* conflict tasks, and so these two debates, one unfolding in cognitive science and the other in social psychology, are likely deeply connected. In what follows, I lay out the rationale for the divergent verdict (ontological vindication/measurement conviction[2]) in the case of conflict tasks, and I argue that we should take a similar position with regard to indirect measures such as the IAT.

## 2 | THE STROOP TASK AND OTHER CONFLICT TASKS

### 2.1 | Conflict tasks provide strong evidence for the *existence* of automatic associations in certain cognitive domains

In the Stroop task, subjects are asked to say the ink color of stimuli that appear on the screen, and these stimuli are themselves color words such as "red" and "green" (Stroop, 1935). On congruent trials, the ink color and color word match, for example, the word "red" is presented in red ink, and on incongruent trials, the stimulus features disagree, for example, the word "red" is presented in green ink. The classic finding, dubbed the Stroop congruency effect, is that response times in incongruent trials are much slower than in congruent trials, and there are differences in other metrics too, for example, accuracy and reaction time variability. The robustness of the Stroop congruency effect is striking. Across studies, the Stroop congruency effect is one of the most widely and consistently replicated results in psychology (MacLeod, 1991). Across individuals within studies, the congruency effect is observed in nearly all subjects (Stroop, 1935). Across time, the effect is temporally stable (Hedge et al., 2018). That is, if you bring subjects back to take the task repeatedly across weeks and months, they will consistently exhibit the Stroop congruency effect across testing sessions.[3]

What causes this congruency effect? The standard explanation, suggested by John Stroop himself in his original 1935 paper (Stroop, 1935), relies on the idea of *automatic associations*. The congruent condition takes advantage of strong automatically activated associations between word stimuli and reading responses, and thus response times are quicker. In the incongruent condition, subjects must produce a less familiar mapping between ink colors and color naming responses, and indeed automatic word reading responses must be actively suppressed, and thus subjects' response times are longer. In short, in the congruent condition, automatic associations are "aligned" with the response, while in the incongruent condition they are "misaligned," resulting in the Stroop congruency effect.

Stroop's "automaticity" hypothesis has across the subsequent years received consistent and convergent support. The evidence for the hypothesis is sizable and scattered across multiple literatures, but some highlights include the following findings. Individuals who are less proficient in English (or whatever language in which the task is conducted) display a correspondingly reduced Stroop congruency effect (Wang et al., 2016), which is best explained by these individuals' having reduced strength of the relevant automatic associations. Non-color words, which do not automatically activate associated color responses, reduce or eliminate the congruency effect (Klein, 1964). Cognitive load increases the congruency effect, which is best explained by load having selective effects on top-down control relative to

activation of automatic associations (Kalanthroff et al., 2015). Perhaps most decisively, computational modeling of the generative processes that produce reaction times and accuracies in the task reveals an early-onset automatic processing component that is clearly distinguished from a later-onset controlled processing component (Ulrich et al., 2015).

There are a wide variety of other conflict tasks as well. For example, the Simon task reveals the presence of automatic associations between lateralized stimuli and spatially concordant responses (Simon & Wolf, 1963). The anti-saccade task reveals automatic associations between movement of visual stimuli and concordant gaze responses (Munoz & Everling, 2004). Taken together, conflict tasks such as these provide strong evidence for the existence of automatic associations in a number of cognitive domains.

## 2.2 | Conflict tasks are relatively poor at *measuring individual differences,* at least with standard analysis approaches

In ordinary language, "measurement" is often used as a synonym for detection. So, one might think that if some phenomenon exists and operates in our minds, then *of course* we can "measure" it. But in psychology and cognitive science, measurement is overwhelmingly used in a technical sense to refer to *measurement of individual differences* (Nunnally, 1970). There are many features that exist and operate in our minds but which we cannot measure—in this individual difference sense—because either these features lack individual differences (they are highly uniform across persons), or the methods we use to measure individual differences are, for whatever reason, ineffective. Indeed, the Stroop task and other conflict tasks provide a clear illustration of just this sort of dissociation between the existence of a construct and our ability to measure individual differences in it.

The standard way to use conflict tasks to study individual differences is somewhat related to the congruency effect noted before. The congruency effect is usually calculated as the average across all subjects of their difference in reaction time between incongruent and congruent conditions. But in assessing individual differences, we do not average across subjects. Instead, each individual is assigned a score corresponding to their own personal difference in reaction time between their incongruent and congruent conditions—thus the name "difference score" (see Figure 1).

For decades, it has been widely assumed by researchers that difference scores from conflict tasks reflect an important individual difference dimension. More specifically, the idea was that some people have more effective top-down control processes, enabling them to more rapidly regulate inappropriately activated automatic associations, thus producing smaller reaction time differences between incongruent versus congruent conditions, while others have less effective control processes yielding larger differences. In short, difference scores in conflict tasks were thought to reflect individual differences in the effectiveness of top-down control.

In the last few years, a number of lines of evidence have emerged that raise serious challenges for this multi-decade orthodoxy. First, recent large studies using multiple, diverse conflict tasks show that difference scores have unacceptably low test–retest reliability (Enkavi et al., 2019; Hedge et al., 2018). Second, difference scores lack cohesion:
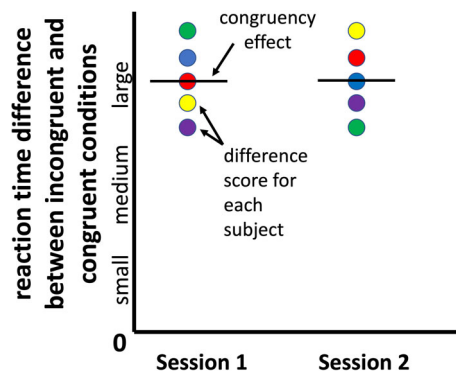


**FIGURE 1** Schematic figure distinguishing temporal stability of congruency effects from test–retest reliability of difference scores. The figure depicts five subjects, each represented by a different color. At session 1, the congruency effect, that is, the average difference between incongruent and congruent reaction times, is large. It is also temporally stable, that is, the effect is the same size across session 1 and session 2. However, subjects' difference scores have very poor test–retest reliability. That is, the rank ordering of subjects by their difference scores is unstable across sessions.

Interrelations among difference scores drawn from conceptually similar conflict tasks are relatively weak (Karr et al., 2018; Weigard et al., 2021), even under conditions in which problems with test–retest reliability are minimized (Rey-Mermet et al., 2019; Rouder & Haaf, 2019). Third, difference scores from conflict tasks have poor predictive validity: Well-powered analyses find difference scores display weak to negligible correlations with the things they are supposed to be related to, such as self-report indices of self-control (Eisenberg et al., 2019; Saunders et al., 2018; Weigard et al., 2021) and psychiatric disorders that involve deficits in self-control (Sripada & Weigard, 2021). These three problems, that is, poor reliability, lack of cohesion, and poor predictive validity, are strikingly similar to the problems that Machery identifies for indirect measures of implicit attitudes, which are discussed in the following section.

It is critical to emphasize that the preceding critique is not of conflict task *congruency effects*, the observation that, averaging across the sample, subjects have markedly longer reaction times in incongruent conditions relative to congruent conditions. They are critiques that are targeted specifically at the use of *difference scores* from conflict tasks to quantify variation across individuals—a very different target (see Figure 1 for clarification). Notice further that in the previous section, conflict task congruency effects were shown to provide solid evidence for the existence of automatic associations in various psychological domains. The present critique of conflict task difference scores does nothing to call this inference into question. So, we have arrived at a divergent verdict. In the previous section, it was shown that conflict tasks provide strong support for the existence of automatic associations, and in this section, it was shown that these tasks fail as measures of individual differences.

Further support for this divergent verdict arises when we consider why conflict tasks do so poorly as individual difference measures. There are two leading hypotheses. The "excess uniformity" hypothesis says that while difference scores in conflict tasks may be large (reaction times in the incongruent condition are typically about one standard deviation larger than in the congruent condition), individuals fail to exhibit meaningful differences in these scores (Hedge et al., 2018). If an effect is relatively uniform across subjects, then even if the effect is very large, it cannot be used to measure individual differences, simply because meaningful inter-individual differences are not present.

The "failure of pure insertion" hypothesis says meaningful individual differences do exist, but the metrics utilized to quantify these differences are inappropriate (Weigard et al., 2021). A basic assumption in the use of difference score metrics from conflict tasks to quantify individual differences is that the congruent and incongruent conditions are matched in all respects but for the operation of the top-down control process, which is engaged in the incongruent condition (to overcome automatic associations) but not the congruent condition (because in this condition, automatic associations favor the correct response and thus do not need to be overcome). There is growing evidence that the pure insertion assumption is false: The top-down control process operates in both the incongruent and congruent condition, but it operates more strongly in the former (to overcome the countervailing automatic associations) (Draheim et al., 2019; Heathcote et al., under review; Weigard et al., 2021). When pure insertion fails, the measurement properties of difference scores correspondingly deteriorate. In effect, the signal related to the measurand is subtracted away through differencing, thus explaining poor psychometric properties of conflict tasks.

While the debate is ongoing, in my view, the second hypothesis is better supported by the evidence (Weigard et al., 2021). But for the present purposes, let us put this debate aside. The point to focus on is that these two explanations for why conflict tasks are poor individual difference measures have *no implications at all* for claims about the existence of automatic associations made in the previous section. Those existence claims from the previous section are based on observations of robust congruency effects in the Stroop task and other conflict tasks. Meanwhile, these two explanations for why difference scores from conflict tasks are poor individual difference measures do not threaten the reality of robust congruency effects. Indeed, each explanation, in its own way, *presupposes* that conflict tasks exhibit robust congruency effects. Thus, the two present explanations for why conflict task difference scores fail as individual difference measures leave the ontological conclusions reached in the previous section fully intact.

## 3 | INDIRECT MEASURES AND IMPLICIT ATTITUDES: APPLYING LESSONS FROM CONFLICT TASKS

The divergent verdict arrived at with conflict tasks is relevant to the debate about indirect measures of implicit attitudes because several indirect measures *are* conflict tasks. This point is readily appreciated by considering the IAT, which, like the Stroop task, includes a congruent condition where stimuli are "aligned" and an incongruent condition where stimuli are "misaligned." In the race IAT specifically, the congruent condition pairs Black faces with negative evaluative

words and White faces with positive evaluative words, while the incongruent condition pairs Black faces with positive evaluative words and White faces with negative evaluative words.

Much like the Stroop task and other conflict tasks, this stimulus manipulation produces very large effects: White American subjects are much slower in the incongruent condition than the congruent condition, and this effect is comparable in magnitude to the Stroop congruency effect (Nosek et al., 2002). Moreover, like the Stroop congruency effect, this race IAT congruency effect is highly robust (among White American task takers). It is consistently replicated across studies (e.g., Greenwald et al., 1998; McConnell & Leibold, 2001; Nosek et al., 2002), and within studies, it is seen in most White American subjects (Nosek et al., 2007). It is also highly temporally stable: When the same subjects take the race IAT across testing sessions separated in time, this congruency effect is consistently observed across sessions (Gawronski et al., 2017). Additionally, strong IAT congruency effects are also seen when the task stimuli pertain to other stigmatized target groups, for example, gay people with heterosexual task takers (Nosek et al., 2007), Muslim individuals with American task takers (Nosek et al., 2007), Turkish individuals with German task takers (Hofmann et al., 2008), and so on.

What explains the presence and pattern of these IAT congruency effects with respect to stigmatized groups? Some researchers have proposed explanations based on familiarity or salience (Rothermund & Wentura, 2001, 2004), but these hypotheses encounter serious problems (Dasgupta et al., 2000; Gawronski, 2002; Greenwald et al., 2005; Ottaway et al., 2001). Overall, the best current explanation for the observed pattern of IAT congruency effects with respect to stigmatized groups mirrors the explanation proposed by John Stroop in 1935 for the Stroop congruency effect: The reaction time advantage for responses that pair people from stigmatized groups with negative evaluative words reflects the operation of automatic associations, in this case associations between mental representations of stigmatized groups and negative evaluative representations.[4] That is, exactly analogous to the Stroop task, when automatic associations are aligned with the needed response (Black-bad, White-good), reaction times are faster, and when they are misaligned (Black-good, White-bad), reaction times are slower.

According to a standard definition, an "implicit attitude" is an automatic association between a social concept and a valenced representation (Fazio, 2007). It follows that, in the same way that conflict tasks provide strong evidence for the existence of automatic associations in various psychological domains, so too *the IAT and other indirect measures provide strong evidence for the existence of valenced automatic associations in various social domains. That is, they provide strong evidence for the existence of implicit attitudes.*

But now that we have clearly separated broader questions about the existence of implicit attiudes from narrow questions about measurement of individual differences, we can ask a further question: Are individual differences in these implicit attitudes measured effectively with the IAT and other indirect measures? The answer is no. Machery elegantly and concisely summarizes the key findings (Machery, 2022a). In particular, he reviews evidence that difference scores from indirect measures have poor test–retest reliability, they lack cohesion across different indirect measures, and they have poor predictive validity—findings that precisely mirror those from recent critiques of difference scores from conflict tasks.

But critically, analogous to conflict tasks, the poor performance of indirect measures as individual difference measures does not threaten the ontological conclusion arrived at earlier. That conclusion that implicit attitudes exist and operate in our minds is derived from the presence and pattern of IAT congruency effects with respect to stigmatized groups. Machery's critique of IAT difference scores as individual difference measures does not cast doubt on the presence or pattern of these congruency effects (see Figure 1), and thus his critique leaves this previous conclusion fully intact.

This point is reinforced when we dig deeper to understand why indirect measures such as the IAT perform poorly for individual difference purposes. Earlier in my discussion of conflict tasks, two explanations were distinguished. I suggested that for conflict tasks, the problem is likely to be "failure of pure insertion." For indirect measures, however, the problem is more likely to be "excess uniformity."[5]

To see why the idea is plausible, consider studies of semantic priming in which subjects first exposed to a prime word, for example, "doctor," are subsequently much faster to recognize words that are semantically related, for example, "nurse" or "hospital" (Meyer & Schvaneveldt, 1971). The standard explanation for semantic priming appeals to automatically activated semantic associations (Anderson, 1983; Collins & Loftus, 1975). But it is widely thought that these semantic associations, inculcated through years of exposure to word co-occurrence patterns during childhood (Landauer & Dumais, 1997), have relatively similar strength across individuals within a linguistic community. Semantic priming nicely illustrates how excess uniformity introduces a particularly striking dissociation between the existence of a construct and our ability to measure individual differences in that construct. Automatic semantic associations exist,

are large, and are remarkably widespread, and yet we cannot measure individual differences because meaningful differences are not there to be measured.

It is possible that the automatic associations directed at stigmatized groups captured in the IAT are similar. That is, these associations are inculcated through years of exposure to stigmatizing stimuli that permeate a cultural milieu, and, as a result, the associations exhibit relatively uniform strength across individuals within a community. Some support for this hypothesis comes from the observation that psychometric properties of the IAT, including predictive validity, improve considerably when making comparisons *across distinct cultural units* (e.g., a northern U.S. state vs. a southern U.S. state) rather than across individuals within a cultural unit (Payne et al., 2017). For our purposes, the key point is that the excess uniformity hypothesis vividly illustrates why a divergent verdict in which implicit attitudes exist but we cannot measure individual differences effectively is eminently reasonable.

## 4 | AVOIDING "ONTOLOGICAL SLIPPAGE"

Supposing the key claims of the previous section are correct, we must be careful to guard against "ontological slippage." This happens when criticisms of the individual difference measurement properties of indirect measures are used to imply a negative verdict on ontological questions, that is, a verdict that implicit attitudes do not exist.

One place where we see ontological slippage is in popular pieces that say that the IAT and other indirect measures are "junk science"[6] or "failed science."[7] These claims are excessively broad and potentially misleading. We typically think that the kinds and relations postulated by a "junk science" are not real; the fact that the relevant kinds and relations do not exist is precisely *why* the science is junk. But, as we have seen, academic researchers' current criticisms of implicit attitudes target the measurement of individual differences in implicit attitudes through tasks such as the IAT. These tasks, however, yield other kinds of evidence, specifically via group-averaged congruency effects, that are not impacted by these criticisms, and which in fact support the existence of implicit attitudes. In this way, these popular "hot takes" are guilty of ontological slippage and mislead readers into strong ontological conclusions that do not follow from the adduced evidence.

Another place slippage arises is in Machery's framing of his argument in terms of "anomalies," which, drawing on Kuhnian philosophy of science, are empirical results that threaten a scientific paradigm and typically lead to its demise (Machery, 2022a, footnote 1). There are two problems with the claim that poor measurement properties of indirect measures constitute anomalies for the implicit attitudes research program.

The first problem is substantive: Are poor individual difference measurement properties of indirect measures truly so stubborn and recalcitrant that we should give up trying to improve them? Machery appears to place substantial weight on the data point that it has been three decades since the advent of these tasks and improvements have not been forthcoming. The case of conflict tasks, however, gives us an alternative perspective. With conflict tasks, problems with their individual difference measurement properties were little noticed for decades and became apparent only recently. Since then, a number of alternative analysis approaches have been proposed (Draheim et al., 2021), including methods based on computational models rather than difference scores (Weigard et al., 2021), and these approaches show promise. We should be cautious of excessive pessimism in the case of indirect measures of implicit attitudes based on three decades of stagnation when in a precisely parallel case involving conflict tasks, stagnation lasted many decades longer and an optimistic resolution may very well be in the offing.

The second problem is more conceptual. Suppose that a divergent verdict is in fact correct: Implicit attitudes against stigmatized groups exist, but we cannot measure individual differences effectively with indirect measures. To make things more concrete, suppose the reason for poor individual difference measurement properties is excess uniformity: People within a group end up with roughly the same level of implicit bias, and thus, meaningful individual differences are not present. It seems very odd to claim that this situation justifies abandoning research into implicit biases. In the parallel case of automatic semantic associations, it is widely thought that these associations are real, strong, and nearly universal among people within a linguistic community, but meaningful individual differences are not present. Far from being abandoned, the study of automatic semantic associations remains a vibrant area of research in psycholinguistics (McNamara, 2005). Presumably this is because the criteria for abandoning a research program is that the kinds and relations it posits fail to exist, not that we cannot measure individual differences in that domain.

It seems, then, that Machery's talk of anomalies may involve some degree of ontological slippage. He is marshaling a case (one that is admirably clear and convincing) that indirect measures are poor at measuring individual differences in implicit attitudes. But to justify abandoning research on implicit attitudes, what is needed is something stronger: an

overall tallying of evidence that shows these attitudes fail to exist. But Machery does not tally such evidence, nor is it even his goal to do so. Indeed, as I have argued at length in this commentary, in a somewhat strange twist, some of the strongest evidence that implicit attitudes do in fact exist comes from the very tasks whose individual difference properties Machery is (correctly) criticizing. In short, talk of anomalies suggests a damning ontological problem for implicit attitudes research, when the actual problem Machery is focused on, measurement of individual differences, is distinct and notably narrower.

## 5 | CONCLUSION

In this commentary, I considered recent criticisms of indirect measures such as the IAT, being careful to separate narrow questions about measurement of individual differences in implicit attitudes from broader questions about whether implicit attitudes exist. Moreover, I drew key insights and lessons from a parallel debate in cognitive science about conflict tasks such as the Stroop task. I argued that the evidence overall supports a nuanced conclusion: Indirect measures such as the IAT measure individual differences in implicit attitudes poorly. At the same time, these tasks—via distinct lines of evidence—still strongly support the view that implicit attitudes exist.

### ENDNOTES

[1] Schimmack (2021) is a notable exception that explicitly focuses their critique on measurement of individual differences. They leave more general questions about the existence of implicit attitudes largely open.

[2] There is one complication to note at the outset: The measurement conviction in the case of conflict tasks appears to be temporary. Evidence is growing for substantially improved measurement properties for these tasks from alternative analysis strategies, which I discuss in more detail later.

[3] Temporal stability of the Stroop congruency effect should not be confused with test–retest reliability of Stroop difference scores. This issue is discussed below and in Figure 1.

[4] This claim leaves open key details about what type of automaticity is at issue. For example, it is left open whether the stimulus-valence association should be understood as a direct causal link or whether it should be understood "propositionally" (Mandelbaum, 2016).

[5] While there are strong parallels between traditional conflict tasks such as the Stroop task and indirect measures of implicit attitudes such as the IAT, there is also a striking difference: In conflict tasks, the intended target of measurement is the *effectiveness of the top-down control process*, while in indirect measures like the IAT, the intended target is the *strength of the automatic association*. This difference in intended measurement target accounts for why there are different explanations for their poor measurement properties, that is, "failure of pure insertion" for conflict tasks and "excess uniformity" for indirect measures. This point deserves a much more detailed treatment, but space does not allow for that here.

[6] www.nationalreview.com/corner/unconscious-bias-roger-clegg/6/

[7] https://leiterreports.typepad.com/blog/2021/06/the-final-nail-in-the-coffin-for-the-failed-science-of-implicit-bias.html

### ORCID
*Chandra Sripada* https://orcid.org/0000-0001-9025-6453

### RELATED WIREs ARTICLE
What Do Implicit Measures Measure

### REFERENCES
Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, *22*(3), 261–295.
Brownstein, M., Madva, A., & Gawronski, B. (2019). What do implicit measures measure? *Wiley Interdisciplinary Reviews: Cognitive Science*, *10*(5), e1501.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428.

Dasgupta, N., McGhee, D. E., Greenwald, A. G., & Banaji, M. R. (2000). Automatic preference for white Americans: Eliminating the familiarity explanation. *Journal of Experimental Social Psychology*, *36*(3), 316–328.

Draheim, C., Mashburn, C. A., Martin, J. D., & Engle, R. W. (2019). Reaction time in differential and developmental research: A review and commentary on the problems and alternatives. *Psychological Bulletin*, *145*(5), 508–535.

Draheim, C., Tsukahara, J. S., Martin, J. D., Mashburn, C. A., & Engle, R. W. (2021). A toolbox approach to improving the measurement of attention control. *Journal of Experimental Psychology: General*, *150*(2), 242–275.

Eisenberg, I. W., Bissett, P. G., Zeynep Enkavi, A., Li, J., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019). Uncovering the structure of self-regulation through data-driven ontology discovery. *Nature Communications*, *10*(1), 1–13.

Enkavi, A. Z., Eisenberg, I. W., Bissett, P. G., Mazza, G. L., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019). Large-scale analysis of test–retest reliabilities of self-regulation measures. *Proceedings of the National Academy of Sciences*, *116*(12), 5472–5477.

Fazio, R. H. (2007). Attitudes as object–evaluation associations of varying strength. *Social Cognition*, *25*(5), 603–637.

Gawronski, B. (2002). What does the implicit association test measure? A test of the convergent and discriminant validity of prejudice-related IATs. *Experimental Psychology*, *49*(3), 171–180.

Gawronski, B., Brownstein, M., & Madva, A. (2022). *How should we think about implicit measures and their empirical 'anomalies'?* WIREs Cognitive Science.

Gawronski, B., Morrison, M., Phills, C. E., & Galdi, S. (2017). Temporal stability of implicit and explicit measures: A longitudinal analysis. *Personality and Social Psychology Bulletin*, *43*(3), 300–312.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480. https://doi.org/10.1037/0022-3514.74.6.1464

Greenwald, A. G., Nosek, B. A., Banaji, M. R., & Christoph Klauer, K. 2005. "Validity of the Salience Asymmetry Interpretation of the Implicit Association Test: Comment on Rothermund and Wentura (2004)."

Heathcote, A., Hannah, K., & Matzke, D. under review. "Pervasive Choice Conflict."

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, *50*(3), 1166–1186.

Hofmann, W., Gschwendner, T., Castelli, L., & Schmitt, M. (2008). Implicit and explicit attitudes and interracial interaction: The moderating role of situationally available control resources. *Group Processes & Intergroup Relations*, *11*(1), 69–87.

Kalanthroff, E., Avnit, A., Henik, A., Davelaar, E. J., & Usher, M. (2015). Stroop proactive control and task conflict are modulated by concurrent working memory load. *Psychonomic Bulletin & Review*, *22*(3), 869–875.

Karr, J. E., Areshenkoff, C. N., Rast, P., Hofer, S. M., Iverson, G. L., & Garcia-Barrera, M. A. (2018). The unity and diversity of executive functions: A systematic review and re-analysis of latent variable studies. *Psychological Bulletin*, *144*(11), 1147–1185.

Klein, G. S. (1964). Semantic power measured through the interference of words with color-naming. *The American Journal of Psychology*, *77*(4), 576–588.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240.

Machery, E. (2022a). Anomalies in implicit attitudes research. *Wiley Interdisciplinary Reviews: Cognitive Science*, *13*(1), e1569.

Machery, E. (2022b). *Anomalies in implicit attitudes research: Not so easily dismissed* (p. e1591). Wiley Interdisciplinary Reviews: Cognitive Science.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109*(2), 163–203.

Mandelbaum, E. (2016). Attitude, inference, association: On the propositional structure of implicit bias. *Noûs*, *50*(3), 629–658.

McConnell, A. R., & Leibold, J. M. (2001). Relations among the implicit association test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology*, *37*(5), 435–442.

McNamara, T. P. (2005). *Semantic priming: Perspectives from memory and word recognition*. Psychology Press.

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*(2), 227–234.

Munoz, D. P., & Everling, S. (2004). Look away: The anti-saccade task and the voluntary control of eye movement. *Nature Reviews Neuroscience*, *5*(3), 218–228.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice*, *6*(1), 101–115.

Nosek, B. A., Smyth, F. L., Hansen, J. J., Devos, T., Lindner, N. M., Ranganath, K. A., Smith, C. T., Olson, K. R., Chugh, D., & Greenwald, A. G. (2007). Pervasiveness and correlates of implicit attitudes and stereotypes. *European Review of Social Psychology*, *18*(1), 36–88.

Nunnally, J. C. (1970). *Introduction to psychological measurement*. McGraw-Hill.

Ottaway, S. A., Hayden, D. C., & Oakes, M. A. (2001). Implicit attitudes and racism: Effects of word familiarity and frequency on the implicit association test. *Social Cognition*, *19*(2), 97–144.

Payne, B. K., Vuletich, H. A., & Lundberg, K. B. (2017). The bias of crowds: How implicit bias bridges personal and systemic prejudice. *Psychological Inquiry*, *28*(4), 233–248.

Rey-Mermet, A., Gade, M., Souza, A. S., Von Bastian, C. C., & Oberauer, K. (2019). Is executive control related to working memory capacity and fluid intelligence? *Journal of Experimental Psychology: General*, *148*(8), 1335–1372.

Rothermund, K., & Wentura, D. (2001). Figure-ground asymmetries in the implicit association test (IAT). *Zeitschrift für Experimentelle Psychologie*, *48*(2), 94–106.

Rothermund, K., & Wentura, D. (2004). Underlying processes in the implicit association test: Dissociating salience from associations. *Journal of Experimental Psychology: General*, *133*(2), 139–165.

Rouder, J. N., & Haaf, J. M. (2019). A psychometrics of individual differences in experimental tasks. *Psychonomic Bulletin & Review*, *26*(2), 452–467.

Saunders, B., Milyavskaya, M., Etz, A., Randles, D., Inzlicht, M., & Vazire, S. (2018). Reported self-control is not meaningfully associated with inhibition-related executive function: A Bayesian analysis. *Collabra: Psychology*, *4*(1), 39.

Schimmack, U. (2021). The implicit association test: A method in search of a construct. *Perspectives on Psychological Science*, *16*(2), 396–414.

Simon, J. R., & Wolf, J. D. (1963). Choice reaction time as a function of angular stimulus-response correspondence and age. *Ergonomics*, *6*(1), 99–105.

Sripada, C., & Weigard, A. (2021). Impaired evidence accumulation as a transdiagnostic vulnerability factor in psychopathology. *Frontiers in Psychiatry*, *12*, 107. https://doi.org/10.3389/fpsyt.2021.627179

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662.

Ulrich, R., Schröter, H., Leuthold, H., & Birngruber, T. (2015). Automatic and controlled stimulus processing in conflict tasks: Superimposed diffusion processes and delta functions. *Cognitive Psychology*, *78*, 148–174.

Wang, R., Fan, X., Liu, C., & Cai, Z. G. (2016). Cognitive control and word recognition speed influence the Stroop effect in bilinguals. *International Journal of Psychology*, *51*(2), 93–101.

Weigard, A., Angus Clark, D., & Sripada, C. (2021). Cognitive efficiency beats top-down control as a reliable individual difference dimension relevant to self-control. *Cognition*, *215*, 104818.