The Forgiveness Implicit Association Test

Jeremy Goldring, Peter Strelan *

School of Psychology, University of Adelaide, North Terrace, Adelaide 5005, Australia

ARTICLE INFO

ABSTRACT

Across nine studies involving $N = 1174$ participants, we report the development and testing of the Forgiveness Implicit Association Test (IAT). We identify appropriate contrast categories and word content (Studies 1–3); address issues related to implicit-explicit convergence (Studies 4 and 5); and test a double dissociation model to examine the conditions under which the Forgiveness IAT predicts transgression-specific forgiveness (Studies 6–9). We also conducted meta-analyses to examine the extent to which the Forgiveness IAT is resistant to socially desirable responding, relative to self-report measures; and the extent to which individuals implicitly prefer forgiveness to several punitive alternatives (e.g., revenge). The Forgiveness IAT appears to be a good complementary measure to existing trait-level self-report forgiveness measures.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

The psychological literature on forgiveness is now very large (for example, the Scopus database currently yields >1300 citations). In the overwhelming majority of cases, forgiveness is measured via self-report. However, there are limitations to measuring forgiveness, in particular, with self-report. First, the reliance on one mode of measurement is in itself problematic. Mono-method bias increases the probability of shared error variance in results (Hoyt & McCullough, 2005). Second, forgiveness is a value-laden construct such that people perceive it as socially desirable or, alternatively, weak and therefore undesirable (see Fisher & Katz, 2000). Either way, asking people to self-report the extent to which they are generally forgiving is susceptible to self-presentation biases: we are generally motivated to present ourselves in a particular light (e.g., Ballard, 1992). Third, individuals do not always possess sufficient self-awareness to provide a realistic portrait of who they are; alternatively, they are not always able to access attitudes or motivations or beliefs of which they are consciously aware (e.g., Greenwald, McGhee, & Schwartz, 1998).

In recent years, researchers have employed an alternative to self-report questionnaires, the Implicit Association Test (IAT; for a review, see Nosek, Greenwald, & Banaji, 2007). In this article, we report the development and application of such a test to forgiveness, focussing on the conditions under which a Forgiveness IAT, relative to traditional self-report measures, will be more or less likely to predict transgression-specific forgiveness.

1.1. The Implicit Association Test

The IAT is a computerised reaction-timed word sorting task, requiring respondents to quickly sort words or pictures into two pairs of categories while making as few errors as possible. The IAT is a relative measure, in which researchers pair the construct of interest (in this case, 'forgiveness') with a contrast category (e.g., 'revenge'). To establish the extent to which a person is more or less inclined towards forgiveness (relative to revenge, for example), the IAT measures the time that participants take to associate opposing attributive categories with forgiveness (or, for example, revenge). The content of attributive categories depends on whether the IAT is attitudinal or self-concept in nature. The attributive categories for an attitudinal IAT are typically 'pleasant/unpleasant' or 'good/bad' (e.g., Greenwald et al., 1998). The attributive categories for a self-concept IAT are usually 'self' and 'others' (Asendorpf, Banse, & Mucke, 2002). Faster performance (after controlling for errors) when *forgiveness* is paired with *pleasant or self* (compared with when *forgiveness* and *unpleasant or other* are paired) is deemed to reflect greater implicit forgiveness at the trait level. Researchers infer dispositional inclinations and attitudinal preferences from task performance (see Nosek et al., 2007 for a review). The term "implicit" reflects the idea that individuals' preference decisions operate outside of conscious awareness. As such, the IAT has the potential to address shortcomings of self-report, particularly in relation to forgiveness.

1.2. Previous research on a Forgiveness IAT

We know of only one previous attempt to develop a Forgiveness IAT (Fatfouta, Schröder-Abé, & Merkl, 2014). Notably, this study found no relation between a Forgiveness IAT and transgression-specific

---

* Corresponding author.
E-mail address: peter.strelan@adelaide.edu.au (P. Strelan).

forgiveness of a recalled transgression, whereas a self-report measure was significantly associated. Further, their IAT seems to have confounded target categories by including first person referential words (i.e., you, yours) as other-oriented words (i.e., they, them). In addition, while they demonstrated that higher scores on their Forgiveness IAT were associated with faster responding to forgiveness items, higher IAT scores were also associated with faster responding to revenge items.

1.3. Overview of studies

In this article, we report the results of nine studies. Studies 1–3 were concerned with IAT development. Studies 4–5 focused on issues related to implicit-explicit convergence (IEC). Studies 6–9 tested the conditions under which the Forgiveness IAT would most likely predict transgression-specific forgiveness. In conducting these nine studies, we also addressed the extent to which the Forgiveness IAT is resistant to socially desirable responding, relative to self-report measures; and the extent to which individuals implicitly prefer forgiveness to several alternatives (e.g., revenge).

For the purpose of economy, we summarize information about the participants in the nine studies in Table 1 (N, M_{age}, sample source, and religiosity).

In Table 2, we summarize the internal reliabilities for the various versions of the Forgiveness IAT employed across the nine studies. Table 2 also summarizes basic details of the self-report measures used across the nine studies. We included six trait-level scales for testing convergent validity (ATF, TTF, HFS, WTF, Vengeance, and FIS), and four scales for examining socially desirable responding. All scales, apart from the FIS, which we developed, have been regularly used in previous research.

2. Study 1

The majority of IATs are attitudinal in nature (Greenwald, Poehlman, Uhlmann, & Banaji, 2009). Thus, we began by developing an attitudinal IAT for forgiveness. Following Greenwald et al. (1998), the attributional categories were *pleasant* (freedom, love, cheer, pleasure, gift, happy), and *unpleasant* (abuse, filth, hatred, poison, evil, tragedy). Because laypeople tend to intuit revenge as the opposite of forgiveness, and revenge is indicated as the opposite of forgiveness in the most widely used transgression-specific measure of forgiveness, McCullough et al.'s (1998) TRIM scale, as a starting point we chose *revenge* as the contrast category to forgiveness. The aim of Study 1 was then to establish appropriate stimulus words for forgiveness and revenge. We were particularly concerned to ensure from the outset that any effects of the Forgiveness IAT were not confounded with target stimuli valence (see Govan & Williams, 2004). That is, respondents could find it easier to associate 'forgiveness' words with 'pleasant' purely based on the valence of these words, rather than as a function of individuals' implicit attitudes.

Thus, in Study 1 we sought to determine if participant strategic recoding based on stimuli valence will be an issue for a Forgiveness

IAT. We therefore varied the valence of stimuli for the forgiveness category, employing three randomly allocated conditions: one in which forgiveness was represented by six positively valanced words (absolve, compassion, mercy, empathy, reconcile, amnesty), one where forgiveness was represented by six negatively valanced words (excuse, overlook, condone, justify, lenient, ignore), and a third "balanced" condition consisting of three positively and three negatively valanced words (absolve, compassion, mercy, excuse, overlook, condone). All words for revenge were negatively valanced (retaliate, vengeance, retribution, payback, vindictive, reprisal). Note that IAT effects are typically unaffected by stimulus set size (see Nosek, 2005).

We selected the stimulus words for forgiveness and revenge from several thesauruses with consideration to the guidelines for stimulus word selection presented in a review of the IAT (see Nosek et al., 2007), as well as examining the results of forgiveness prototype analyses (e.g., Kearns & Fincham, 2004).

The Forgiveness-revenge attitudinal IAT followed standard procedures as outlined by Greenwald et al., (1998), with the exception of the number of trials per block. The present study used blocks of 24 and 48, rather than 20 and 40 respectively, as this accommodated the number of word exemplars used in the IATs—in this case, six per category, ensuring each exemplar was shown an equal number of times. The sequence and category pairings of the IAT trials are outlined in Table 3.

Internal consistency reliability for IATs is usually calculated based on the log-transformed differences in responses latencies on corresponding trials between the compatible (blocks 3 and 4) and incompatible (blocks 6 and 7) blocks, e.g. the first trial of block 6 minus the first trial of block 3, the sixteenth trial on block 7 minus the sixteenth trial on block 4. These difference (D) scores are then treated as individual scale items, and Cronbach's alpha calculated for them accordingly (for a detailed explanation of these steps see Egloff & Schmukle, 2003). Table 2 summarizes the Forgiveness IAT alphas across the nine studies.

2.1. Results

We calculated IAT D scores following the revised scoring algorithm outlined by Greenwald, Nosek, and Banaji (2003), which includes steps to address extreme fast responses (<300 ms), as well as taking into account the number of errors made by participants. Participants for whom >10% of responses were faster than 300 ms (N = 20) were excluded from the calculations and subsequent analyses, leaving 136 participants. Before calculating D scores, response latencies for each error trial were replaced with the mean latency for that block plus a 600 ms penalty. We then calculated D scores by subtracting means latencies for blocks 3 and 4 from blocks 6 and 7, and then dividing by the pooled standard deviation of these four blocks.

D scores have a possible range between −2 and +2, with a score of zero indicating no/equal preference for the two target constructs. In this study, the categories were coded such that positive D scores indicated an implicit preference for forgiveness relative to revenge, whereas negative D scores reflected an implicit preference for revenge relative to

Table 1
Summary of participant information for all studies.

Study	N (F)	M_{age} (SD)	Sample	% Christian/Buddhist/Muslim/Jewish/Hindu/other/Non-religious
1	156 (103)	20 (4.17)	Undergraduates	43/10/1/1/1/6/38
2	114 (87)	19 (3.72)	Undergraduates	55/10/2/2/1/9/22
3	215 (145)	21 (5.76)	Undergraduates	41/9/1/1/1/8/39
4	217 (144)	25 (11.10)	Community sample Facebook	41/11/2/1/1/9/35
5	224 (193)	33 (12.54)	Community sample Facebook	49/9/0/2/1/14/25
6	86 (56)	23 (5.45)	Undergraduates	Not measured
7	147 (94)	21 (4.90)	Undergraduates	49/6/2/1/1/6/35
8	75 (51)	20 (4.33)	Undergraduates	59/0/9/0/3/5/24
9	70 (55)	20 (4.10)	Undergraduates	49/10/3/1/0/6/31

Note. All studies were conducted online apart from Study 6, which was conducted in a laboratory. All participants were Australian residents.