THE IMPLICIT RELATIONAL ASSESSMENT PROCEDURE (IRAP) AND THE MALLEABILITY OF AGEIST ATTITUDES

Claire Cullen, Dermot Barnes-Holmes, Yvonne Barnes-Holmes

National University of Ireland, Maynooth

Ian Stewart

National University of Ireland, Galway

The current study examined the malleability of implicit attitudes using the Implicit Relational Assessment Procedure (IRAP). In Experiment 1, "similar" and "opposite" were presented as response options with the sample terms "old people" and "young people" and various positive and negative target stimuli. Results showed significantly faster response latencies for consistent (e.g., Similar-Positive-Young People) compared to inconsistent tasks (e.g., Similar-Positive-Old People). Explicit measures did not correlate with this IRAP effect. Experiment 2 determined whether prior exposure to pictures of admired and disliked old and young individuals had an impact on IRAP performance. Results revealed that pro-old exemplars reduced the pro-young IRAP effect, but reversed the anti-old effect, and this held for 24 h; explicit measures were largely unaffected. The findings suggest that the IRAP provides an informative measure of attitude-change following pro- versus anti-exemplar training.

Keywords: implicit attitudes, transformation of function, multiple exemplars, implicit cognition, relational frame theory, adult humans

The study of implicit cognition has led to a recent profusion of research within psychology that has served to generate a range of so-called implicit measures, such as the *Implicit Association Test* (IAT; Greenwald, McGhee, & Schwartz, 1998), the Go/No-go Association Task (GNAT; Nosek & Banaji, 2001), Evaluative Priming (Fazio, Sanbonmatsu, Powell, & Kardes, 1986), and the Extrinsic Affective Simon Task (EAST; De Houwer, 2003). Currently, the most popular among these measures is the IAT, and its basic effect has been replicated many times (Greenwald, Nosek, Banaji, & Klauer, 2005). For

example, the IAT has been used to assess implicit cognitions in the socially sensitive domains of homophobia (e.g., Banse, Seise, & Zerbes, 2001), gender stereotypes (e.g., Rudman & Glick, 2001), racism (e.g., Greenwald, et al., 1998), religious stereotyping (e.g., Rudman, Greenwald, Mellott, & Schwartz, 1999), and even self-esteem (e.g., Bosson, Swann, & Pennebaker, 2000).

The IAT, as an indirect measure of implicit attitudes, rests on the assumption that it should be easier to categorize concepts that are strongly associated in memory relative to concepts that are weakly associated (Greenwald et al., 1998). In a seminal study, Greenwald et al. (Experiment 1) used the IAT to test responses to four categories of items (flowers, insects, pleasant words, and unpleasant words). The researchers assumed that the concept *flower* and the attribute *pleasant* are associated in memory, as are the concept *insect* and the attribute *unpleasant.* Based on this assumption, Greenwald et al. reasoned that responses should be faster when response key assignment was congruent (e.g., key 1 = "flower" and "pleasant" vs. key 2 = "insect" and "unpleasant"), rather than incongruent (e.g., key 1 = "flower" and "unpleasant" vs. key 2 = "insect" and "pleasant"). As predicted, mean response latencies were shorter for congruent relative to incongruent tasks. Critically, a divergence between performance on the IAT and the expression of explicit attitudes has been repeatedly shown (e.g., Greenwald & Banaji, 1995; De Houwer, 2002; de Jong, 2002). For example, participants who alleged *not* to hold prejudiced attitudes responded faster and more accurately when categorizing White people with positive words and Black people with negative words, as opposed to White with negative and Black with positive (see Greenwald et al., 2002).

Traditionally, research on prejudice reduction has viewed implicit attitudes as fixed and immutable and has therefore focused on changing explicit attitudes (Bargh, 1999; Devine, 1989). In the past, it was argued that prejudice reduction requires perceivers to be aware of their bias and be motivated to change it (Allport, 1954; Devine, Monteith, Zuwerink, & Elliot, 1991; Myrdal, 1944). However, a recent theme in the literature concerns the malleability of implicit cognitions, which are believed to be less available to introspection and deliberate control (see Bargh, 1999; Blair, 2002). Several recent studies have suggested that implicit attitudes are sensitive to (a) expectancies (Blair & Banaji, 1996), (b) practice or training (Kawakami, Dovidio, Moll, Hermsen, & Russin, 2000), (c) automatic motives (Moskowitz, Salomon, & Taylor, 2000), and (d) motivation to respond without prejudice (Lepore & Brown, 1997).

One way in which researchers have begun to study the malleability of implicit attitudes involves presenting participants with a series of exemplars that are designed to affect their attitudes toward a specific target (e.g., Dasgupta & Greenwald, 2001; Lowery, Hardin, & Sinclair, 2001). In Experiment 1 of the study reported by Dasgupta and Greenwald, participants were exposed to pictures of either admired Black and disliked White individuals, or disliked Black and admired White individuals. An IAT was administered directly after exemplar exposure and again 24 h later without reexposure to the exemplars. Explicit attitude measures were also administered across the two sessions. Results revealed that exposure to admired Black and disliked White exemplars significantly weakened implicit pro-White preferences for 24 h, but did not affect explicit racial attitudes. Experiment 2 replicated Experiment 1 but focused on implicit age-related attitudes. Specifically, participants were exposed to either pro-old (i.e.,

admired old and disliked young individuals) or pro-young (i.e., disliked old and admired young individuals) exemplars. Again, results revealed that exposure to pro-old exemplars significantly weakened implicit pro-young preferences, but not explicit preferences.

Although the IAT has been used with relative success to study implicit attitudes and their malleability, a number of limitations inherent in the procedure have been identified (see De Houwer, 2002). The most important of these in the context of the current study is that the IAT provides a measure of *relative* associative strength and cannot be used to measure the valence of individual concepts (De Houwer, 2002; Nosek, Greenwald, & Banaji, 2004). For example, an IAT effect for "young" versus "old" could indicate that a participant has a positive attitude toward "young" and a neutral attitude toward "old," or it could indicate a neutral attitude toward "young" and a negative attitude toward "old." In effect, the IAT can indicate that *x* is preferred to *y*, but it cannot reveal to what extent *x* and *y* are liked or disliked, per se.

Critically, the relative nature of the IAT effect may serve to obfuscate exactly how a specific attitude-changing intervention affects the targeted attitude. Consider, for example, the study by Dasgupta and Greenwald (2001). Following pro-old and anti-young exemplar training, participants showed a reduced IAT effect. However, did this reduction reflect a reduction in anti-old attitudes, a reduction in pro-young attitudes, or some combination of both? A single IAT cannot answer this question.

Alternative methods for assessing implicit cognition have been developed that aim to assess implicit attitudes toward individual concepts. For example, Nosek and Banaji's (2001) Go/No-Go Association Task (GNAT) is similar to the IAT in the sense that it involves two separate tasks during which participants are required to categorize attribute and target stimuli. Unlike the IAT, however, participants do not have to respond to all of the stimuli on each trial. To assess attitudes toward fruit, for instance, participants may be required in one task to press a key in the presence of fruit names and positive words, and in a second task to press the key when presented with fruit names and negative words. The strength of association between fruit + good versus fruit + bad is thought to be reflected in more rapid responding to fruit when it is paired with good rather than bad. Other tasks, such as the emotional Stroop (Pratto & John, 1991), evaluative priming (Fazio et al., 1986), and the Extrinsic Affective Simon Task (EAST; De Houwer, 2003), have also been offered as methods for assessing implicit attitudes in a nonrelative fashion (see Blair, 2002; De Houwer, 2003). Critically, however, each of these methods, along with the IAT, may be considered a relatively indirect measure of implicit attitudes or beliefs. That is, none of the methodologies requires that participants engage in a task that asks them to confirm or deny, in a direct way, the belief or attitude under investigation.

In noting the indirect nature of the IAT, De Houwer (2002) argued that

> [it] does not provide a measure of beliefs, nor was it designed to do so. It can only provide an index of associations that are assumed to be involved in certain beliefs and thus indirect evidence for the presence of certain beliefs. (pp. 117–118)

In other words, if a methodology such as the IAT indicates that "old"-related and negative words are strongly associated, for example, it is then inferred that such implicit associations underlie negative beliefs about old

people. Although such an inference seems reasonable, it would also seem prudent to attempt to develop additional methodologies that aim to provide relatively direct measures of implicit cognition. One such methodology has recently been offered, the Implicit Relational Assessment Procedure (IRAP; Barnes-Holmes, Barnes-Holmes, Milne, Power, & Stewart, 2006), and, critically, as will be explained later, its structure may also permit a more fine-grained assessment of the impact of exemplar training than that afforded by the IAT.

Unlike the IAT and other indirect measures, each trial of the IRAP asks participants to confirm or deny a specific belief or attitude directly, by responding to a relation between a sample stimulus and a target term.[1] In the first IRAP study (Barnes-Holmes, Hayden, Barnes-Holmes, & Stewart, 2008), participants were presented with a sample stimulus (i.e., "Pleasant" or "Unpleasant"), a positively or negatively valenced target stimulus (e.g., love or an accident), and two relational terms (i.e., *similar* and *opposite*). The feedback for consistent and inconsistent blocks of trials coordinated with and opposed, respectively, previously established verbal relations. As predicted, response latencies were faster for consistent than for inconsistent trials (e.g., participants responded more quickly to Pleasant–Love–Similar than to Pleasant–Vomit–Similar). This basic IRAP effect has now been replicated across a small number of studies, which have shown that the IRAP (a) compares well with the IAT as a measure of individual differences (Barnes-Holmes, Murtagh, Barnes-Holmes, & Stewart, in press; Barnes-Holmes, Waldron, Barnes-Holmes, & Stewart, 2009), (b) is not easily faked (McKenna, Barnes-Holmes, Barnes-Holmes, & Stewart, 2007), (c) may be used as a measure of implicit self-esteem (Vahey, Barnes-Holmes, Barnes-Holmes, & Stewart, 2009), and (d) produces effects that clearly diverge from those obtained from explicit measures when targeting socially sensitive attitudes (Power, Barnes-Holmes, Barnes-Holmes, & Stewart, in press).

Conceptually, the IRAP was derived from Relational Frame Theory (RFT), a behavior analytic account of human language and cognition (Hayes, Barnes-Holmes, & Roche, 2001). According to RFT, the IRAP is a computer-based task that requires participants to respond quickly and accurately in ways that are either consistent or inconsistent with their prior verbal learning histories. It is assumed that overt relational responses defined as consistent on the IRAP will be preceded by incipient or private responses that occur at a higher probability than those responses defined as inconsistent; the probability of such responses is assumed to be determined

---

1   We have suggested that the directness of a measure may lie on a continuum, with the IAT (and its derivatives) at the indirect end, traditional questionnaires at the opposite end, and the IRAP located typically somewhere in the middle (Barnes-Holmes, et al., 2006). Our basic argument is that questionnaires are best considered highly direct measures because they ask participants to respond to items concerning the relevant attitude, and those responses are then assumed to provide a direct measure of same. In contrast, measures such as the IAT are considered indirect because they do not ask participants to respond to questions directly relevant to the attitude under study. Furthermore, differential response times are used to assess associative strength, which is then used to infer the strength of the relevant attitude. The IRAP appears to lie in between these two extremes because it involves asking participants to respond to "questions" directly relevant to the attitude under study, but differential response time is used to index the strength of that attitude. Although this interpretation of the directness of various measures remains open to further debate, it does seem to provide a useful way of conceptualizing the various methodologies.

by historical and current contextual variables. The basic rationale behind the IRAP is that participants' responding should be faster on consistent relative to inconsistent trials because incipient relational responding will coordinate more frequently with the consistent overt responding. In other words, during inconsistent trials, participants' responding is expected to be slower, because they respond against their more probable incipient relational responses.[2]

Given our assumption that historical and current contextual variables determine performance on an IRAP, it seems likely that the types of exemplar exposure employed by Dasgupta and Greenwald (2001) with the IAT should also affect the IRAP. Imagine, for example, that participants are exposed to an IRAP in which they are required to relate positive and negative words, such as "happy" versus "weary," respectively, to the sample stimuli "young people" and "old people." Given the historical context provided by the wider verbal community, prevalent in the Western world, faster responding on the IRAP may be predicted for responses that coordinate *young* with positive and *old* with negative stimuli, than vice versa. If, however, participants are presented with a series of exemplars of positive old people and negative young people, this may establish the IRAP experiment as a context for the transformation of functions of the sample stimulus "old people" from a negative to positive valence, with the opposite effect for the "young people" sample (i.e., from positive to negative). In short, in the absence of any exemplar exposure, the IRAP should produce a pro-young/anti-old effect, but if pro-old/anti-young exemplars are presented, evidence for pro-old and/or anti-young responding may be observed. A primary purpose of the current study was to test this prediction.

As noted earlier, an important feature of the IRAP is that its structure may permit a more fine-grained analysis of the impact of exemplar training than that afforded by the IAT. More specifically, the IRAP consists of four trial types (see Figure 1), and this structure may allow the researcher to assess specific, rather than relative, changes in implicit responses. In the present research, each trial on the IRAP involved presenting one of two sample stimuli—"young people" or "old people"—and a target stimulus, which was either a negative stereotypical term for old people (e.g., "slow," "tired," "stagnant") or a positive term for young people (e.g., "enthusiastic," "energetic," "creative"). The fact that the two categories ("old" and "young") are presented on separate trials with just one target may permit a less relativistic measure of implicit responses to those categories than that provided by the IAT (which presents both categories on every trial during the critical test blocks). For example, if participants are presented with pro-old/anti-young exemplars before exposure to the IRAP, will the IRAP effect for both young and old trial types change, or will we see a change for one but not the other? The current research addressed this issue.

The IRAP is relatively new and it has not been used to examine age-related bias. Thus, Experiment 1 was designed to determine if an IRAP indeed produces a pro-young and/or anti-old bias in the absence of any exemplar exposure. Experiment 2 then sought to investigate the malleability

---

2 We assume that responding against incipient relational responses may occur at an unconscious or nondeliberative level (Barnes-Holmes et al., 2006). This view is broadly consistent with recent evidence indicating that when participants are successfully motivated to reduce an IAT effect, it appears that they do so through unconscious processes (Boysen, Vogel, & Madon, 2006).

of responding on the age-related IRAP using the type of exemplar training employed by Dasgupta and Greenwald (2001).
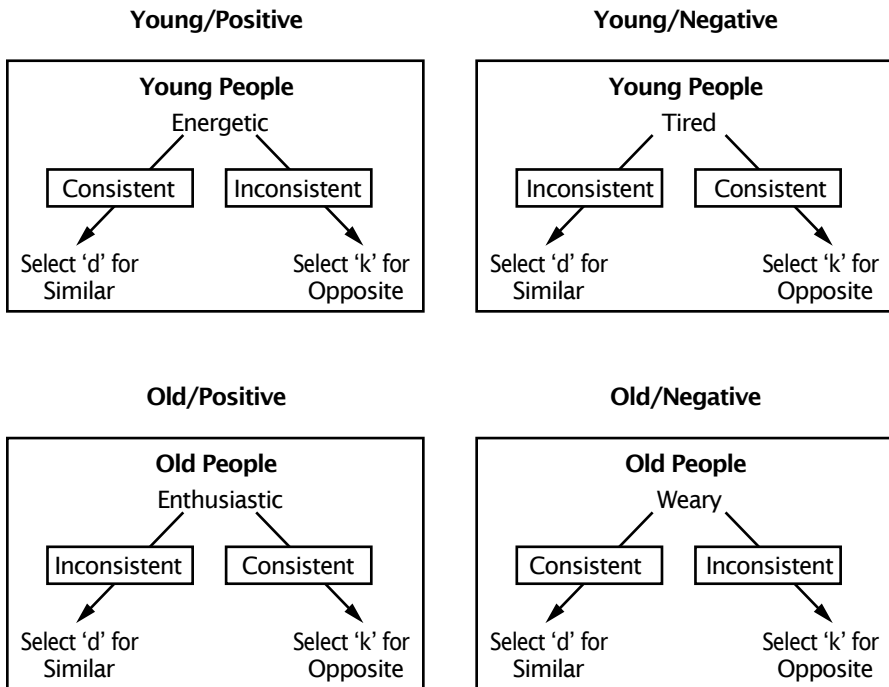
Young/Positive

| Young People |
| Energetic |

Consistent   Inconsistent

Select 'd' for   Select 'k' for
Similar   Opposite

Young/Negative

| Young People |
| Tired |

Inconsistent   Consistent

Select 'd' for   Select 'k' for
Similar   Opposite

Old/Positive

| Old People |
| Enthusiastic |

Inconsistent   Consistent

Select 'd' for   Select 'k' for
Similar   Opposite

Old/Negative

| Old People |
| Weary |

Consistent   Inconsistent

Select 'd' for   Select 'k' for
Similar   Opposite

Figure 1. Examples of the four IRAP trial types. The sample ("young people" or "old people"), target word (*energetic*, *tired*, etc.), and response options (Similar and Opposite) appeared simultaneously on each trial. Arrows with superimposed text boxes indicate which responses were deemed consistent or inconsistent (boxes and arrows did not appear on-screen). Selecting the consistent response option during a consistent block, or the inconsistent option during an inconsistent block, cleared the screen for 400 ms before the next trial was presented; if the inconsistent option was chosen during a consistent block, or the consistent option during an inconsistent block, a red X appeared on-screen until the participant emitted the alternative response.

Experiment 1

Method

Participants

Five males and seven females ranging in age from 18 to 26 years (*M* = 21) completed Experiment 1. All 12 participants were Irish nationals. Ten participants were undergraduate Arts students attending various third-level Irish educational institutes, and two worked as hairdressers. No financial or other incentives, other than the knowledge that they were assisting in scientific research, were offered for participation in the experiment.

Apparatus

Implicit Relational Assessment Procedure. The IRAP software presented stimuli and recorded participants' responses. On each trial, one of two target stimuli ("old people" or "young people") and a single positive or negative target stimulus were presented by the program (see Figure 1 and Table 1). Two response options (i.e. "similar" and "opposite") were also presented on each IRAP trial. The positive and negative target stimuli were selected following a "brainstorming" session in which the investigators sought to generate a range of positive descriptors for young people and negative descriptors for old people. In constructing the list for old people, it was deemed important to identify terms that were negative but not extremely so (e.g., "tired" rather than "evil"), and that in principle could also apply to young people. The converse was true for young people, in that the terms chosen were positive and could in principle apply to old people.

Semantic differentials and feeling thermometers. Participants completed 12 semantic differential scales: 6 for old people and 6 for young people. These 7-point scales (–3 to +3) were anchored at either end by the following polar-opposite adjective pairs (taken from the IRAP): *creative–stagnant*, *happy–sad*, *brilliant–slow*, *energetic–tired*, *enthusiastic–weary*, and *productive–finished*. Two feeling thermometers assessed the favorability of participants' explicit feelings about young and old people. Participants were asked to mark an appropriate position on a picture of a thermometer numerically labelled at 10º intervals from 0º (cold, or unfavorable) to 99º (warm, or favorable). Both the semantic differential scales and feeling thermometers were presented on A4-sized paper sheets.

Table 1
The Stimulus Arrangements for All Trial-types

Sample 1: Young People	Sample 2: Old People
Response Option 1: Similar	Response Option 2: Opposite
Target Stimuli Consistent with Sample 1	Target Stimuli Consistent with Sample 2
Happy	Sad
Energetic	Tired
Productive	Finished
Brilliant	Slow
Creative	Stagnant
Enthusiastic	Weary

Procedure

The experiment consisted of two phases. Phase 1 involved exposure to the IRAP; six participants each were randomly assigned to one of two IRAP conditions (described subsequently). In Phase 2 participants completed the semantic differential scales and feeling thermometers. Each participant completed the experiment on an individual basis, in a quiet room, with the door closed.

Implicit measure. Included in the IRAP computer program were on-screen standardized instructions, which participants read in their own time, pressing the space bar to move between screens. The instructions described the IRAP procedure, described how to complete the task, and

emphasized the need for both accuracy and speed. At no point, however, were the participants informed which tasks were deemed to be consistent or inconsistent.

On each trial of the IRAP, four stimulus words appeared on-screen simultaneously. Specifically, for each trial a sample stimulus, either "old people" or "young people," appeared at the top-center position of the screen with a single target word (e.g., *enthusiastic*) that appeared in the mid-center of the screen. The two relational terms *similar* and *opposite* appeared at the bottom left- and right-hand corners. The phrases "PRESS 'd' FOR" and "PRESS 'k' FOR" appeared directly above the two relational terms. The relational terms alternated position randomly from left to right across trials.

Participants were required to choose one of the two relational terms by pressing the appropriate response key when presented with a sample and target stimulus; all other computer keys were disabled. Choosing the relational term that was deemed "correct" for that particular block of trials removed all stimuli from the screen for a 400-ms interval before the next trial was presented. Choosing the relational term that was deemed "incorrect" for that particular block of trials produced a red "X" mid-screen directly below the target stimulus. The IRAP program only proceeded to the 400-ms interval (and the next trial) when the "correct" relational term was selected.

The IRAP comprised eight blocks of 24 trials. Specifically, two practice blocks were followed by six test blocks. Within each block, the two sample stimuli ("old people" and "young people") were presented randomly across trials, with the constraint that each was presented 12 times within a 24-trial block. The 12 target stimuli were also presented in a random sequence, with the added constraint that each term was presented twice, once in the presence of each sample stimulus.

Two IRAP sequences were employed in the experiment. Six participants were exposed to consistent relations first, and the remaining six were exposed to inconsistent relations first. The first block of trials for those in the former sequence reinforced responses that were deemed relationally consistent with a pro-young/anti-old bias (hereafter referred to as "pro-young"). For example, when presented with the sample "young people" and any of the positive target words, choosing the relational term *similar* was deemed correct and progressed the IRAP program to the next trial. If the relational term *opposite* was chosen, this was deemed incorrect and the red "X" appeared. The same feedback contingencies were applied to the other three trial-types: "young people" – Negative Target – Opposite = Correct / Similar = Incorrect; "old people" – Positive Target – Similar = Incorrect / Opposite = Correct; "old people" – Negative Target – Similar = Correct / Opposite = Incorrect. The second block of trials in the consistent-relations-first sequence reinforced responses that were deemed relationally inconsistent with a pro-young bias (i.e., each of the feedback contingencies described above was reversed; see Figure 1). Following the first two practice blocks, six test blocks were presented in a sequence that alternated between consistent and inconsistent blocks. The same procedures applied in the inconsistent-relations-first sequence, except that the blocks alternated from inconsistent to consistent across both the practice and test blocks.

Before each block of trials, a message appeared on-screen informing the participant that the following block was either a practice or a test. In the latter case, the message also stated, "Go fast—a few errors are okay." Following each

block of trials, feedback was presented on-screen detailing the percentage of correct responses and the median response latency for that block. In addition, a message informed participants that all of the previously correct and wrong answers would be reversed in the next block. Upon completion of all eight blocks, a message appeared asking the participant to report to the researcher.

Explicit measures. Participants completed the six semantic differential scales for old people and the other six for young people. The participants also completed the two feeling thermometers. Participants then reported to the experimenter, who thanked and debriefed them.

Results and Discussion

Data Preparation

The primary datum from the IRAP was response latency, defined as the time in milliseconds that elapsed between the onset of the trial and a correct response made by the participant. For the purposes of the current study, the Greenwald, Nosek, and Banaji (2003) C4 algorithm was applied to the data from the six test blocks. Specifically, latencies below 300 ms and above 3,000 ms were recoded as 300 ms and 3,000 ms, respectively; latencies on error trials were included in the analysis.

Implicit Measure

The overall mean latencies calculated across participants for each of the test blocks for consistent and inconsistent trials are presented in Table 2. Each of the test blocks produced shorter latencies for consistent relative to inconsistent trials. Test blocks 2 and 3 produced shorter latencies relative to test block 1, in both consistent and inconsistent conditions, suggesting a standard practice effect.

Table 2
Overall Mean Response Latencies (in Milliseconds) and Standard Errors in Each Test Block for Consistent and Inconsistent Trials, Separated According to IRAP Sequence; Consistent First Versus Inconsistent First

	Consistent						Inconsistent					
	Test 1		Test 2		Test 3		Test 1		Test 2		Test 3	
	M	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Con 1st	1,909	128	1,787	51	1,774	53	2,069	168	1,985	167	1,930	146
Incon 1st	1,539	80	1,496	67	1,521	137	1,733	124	1,580	103	1,639	84

Preliminary statistical analyses were conducted to test for possible order and sequence effects and to determine if there was an interaction between or among these two variables and the critical IRAP condition (i.e., consistent vs. inconsistent trials). The adjusted mean latencies for each participant were subjected to a $2 \times 2 \times 3$ mixed repeated measures ANOVA, with IRAP conditions and test sequence (Test 1, 2, or 3) as repeated measures and order (consistent-first vs. inconsistent-first) as the between-participants variable.

The main effect for the IRAP condition was significant, $F(1, 10) = 6.51$, $p = .03$, $\eta_p^2 = .39$,[3] as was the effect for order, $F(1, 10) = 5.85$, $p = .04$, $\eta_p^2 = .37$. The main effect for test sequence was nonsignificant, as were the four interaction effects (all $ps > .1$). The analyses indicate that participants responded faster on IRAP trials that involved categorizing young people positively and old people negatively relative to trials that involved the opposite categorizations (young-negative and old-positive; see Figure 2). The main effect for order indicates that participants responded more rapidly when exposed to inconsistent-first relative to consistent-first test blocks, but, critically, neither this variable nor test sequence interacted significantly with the IRAP condition. In short, the predicted consistent versus inconsistent IRAP effect was observed for age-related items, and this effect was not significantly modulated by order and test sequence variables.
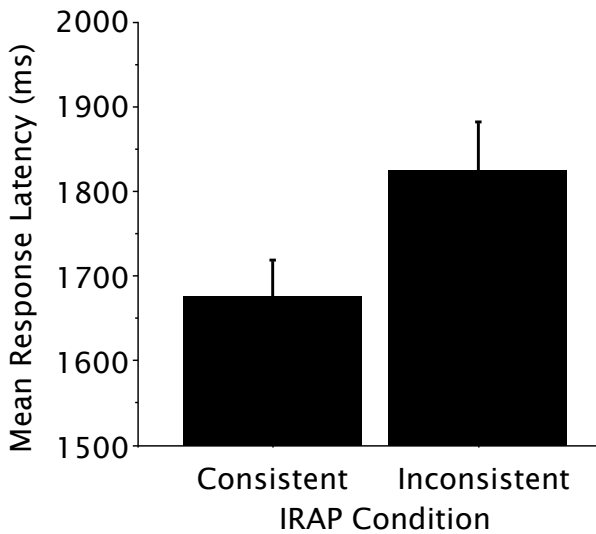

Figure 2. Overall mean response latencies (in milliseconds), including standard error bars, for consistent and inconsistent test blocks.

Explicit Measures

Semantic differential scales were scored by averaging the six items for *old* and then the six items for *young* to create two indices, in which higher scores indicated more favorable attitudes. The two feeling thermometers yielded an evaluative rating for each age category in which higher scores also represented more favorable attitudes. Participants expressed a greater liking for young compared with old people on semantic differential scales ($Ms = 1.94$ and -0.31; $d = 1.68$), and a one-way repeated measures ANOVA indicated that this difference was significant, $F(1, 11) = 16.489$, $p = .002$, $\eta_p^2 = .6$. Participants also reported slightly more favorable attitudes toward young compared with old people on the feeling thermometers ($Ms = 66.67°$

---

3   The partial eta squared (η_p^2) statistic indicates the proportion of the variance (controlling for other variables within the experiment) that is attributable to the relevant effect. This measure of *effect size* is robust at values at or above 0.15 (Cohen, 1988).

and 70.83°; d = .32), but the difference was not significant, $F(1, 11)$ = 1.211, p = .29.

Relationship Between Implicit and Explicit Measures

An IRAP difference score was calculated for each participant by subtracting the total mean response latency for consistent trials from the total mean latency for inconsistent trials; a positive score thus indicated a bias for young people. For explicit measures (both semantic differentials and thermometers), difference scores were computed by subtracting participants' ratings of "old people" from their ratings of "young people," such that higher scores indicated relative preference for young over old.

No significant correlations were found between the explicit and implicit measures (IRAP and feeling thermometers, r = –0.182, p = .55; IRAP and semantic differentials, r = –0.369, p = .22). Similarly, there were no significant correlations between the two explicit measures for ratings of either old or young people (old people, r = 0.234, p = .44; young people, r = 0.252, p = .4).

Summary

The IRAP showed a bias toward young people, as did the semantic differential scales. The IRAP thus performed like the IAT, but the semantic differentials produced an effect not reported by Dasgupta and Greenwald (2001). One likely explanation for this discrepancy is that the current semantic differentials presented relatively mild and more age-relevant descriptors (e.g., *tired–energetic*) compared to the IAT study, which presented such descriptors as *ugly–beautiful, bad–good, unpleasant–pleasant, dishonest–honest,* and *awful–nice.* It seems likely, therefore, that the semantic differentials in the current study were less prone to social desirability effects (e.g., reporting that young people are more energetic than old people might not be seen as particularly prejudiced). The correlations between implicit and explicit measures were all weak and nonsignificant. In any case, the IRAP appears to function in a way that is broadly similar to the IAT. Having demonstrated that the IRAP is sensitive to a pro-young bias, Experiment 2 sought to replicate the Dasgupta and Greenwald (2001) malleability finding, but using the IRAP to assess the category-specific changes that might occur following pro- versus anti-exemplar training.

Experiment 2

Method

Participants

Twenty-four participants (8 males and 16 females) ranging in age from 18 to 26 years (M = 19) completed the first session of the experiment (1 participant failed to return for a second session). Experimentally naïve participants were recruited using the same sample and inclusion criteria as in Experiment 1. Before signing up, however, participants were informed that the study required two half-hour sessions separated by 24 h.

Apparatus

The apparatus and materials employed in Experiment 1 were also employed in Experiment 2, but additional materials were used for the exemplar training (see below).

Selection of exemplars. Pictures of 40 well-known old and young individuals were downloaded from the Internet using the Google™ search engine. Ten pictures featured each of the following four categories: (a) admired old individuals (e.g., Einstein), (b) admired young individuals (e.g., actress Jodie Foster), (c) disliked old individuals (e.g., Imelda Marcos), and (d) disliked young individuals (e.g., Timothy McVeigh). All pictures were converted into gray scale and standardized to 4 cm × 5 cm in dimension. Each picture was then centered on a single sheet of paper, and these were used to construct six separate booklets: two practice booklets, two test booklets, and two age-test booklets.

The practice booklet for the pro-old condition contained 10 pictures of admired old and 10 pictures of disliked young individuals. The pro-young booklet contained 10 pictures of admired young and 10 of disliked old individuals. The name of each person was positioned 1 cm above the picture. Just below the picture were two profile descriptions placed side by side, one being true and the other one false. Throughout the booklet, the left–right positions of these true and false descriptions alternated randomly. An "answer box" appeared directly beneath these descriptions with the words *A or B?* printed underneath. Printed on the back of each page was the correct profile.

The test booklet for each condition was identical to the practice booklet, minus the answers on the back of each page. The two age-test booklets were similar to the test booklets, but there were no pictures or profiles; only the name of each individual and the words *Old* and *Young* were presented, with *A or B* and the answer box.

Implicit Measure

The IRAP was identical to Experiment 1.

Explicit Measures

The six semantic differential scales and two feeling thermometers that were used in Experiment 1 were employed again.

Procedure

Twenty-four participants each were assigned randomly to one of two conditions: 12 to a pro-old/anti-young condition (hereafter referred to as *pro-old*) and 12 to a pro-young/anti-old condition (hereafter referred to as *pro-young*). In each case, 6 participants were assigned to the consistent-relations-first sequence and 6 to the inconsistent–relations-first sequence. The experiment consisted of five phases: (a) exemplar exposure, (b) implicit measure, (c) explicit measures, (d) Implicit measure administered 24 h later, and (e) explicit measures administered 24 h later.

Exemplar exposure. The exemplar task was presented as a "general knowledge task" assessing participants' familiarity with famous and infamous indi-

Table 3
Overall Mean Response Latencies (in Milliseconds) and Standard Errors in Each Test Block for Consistent and Inconsistent Trials, Separated According to IRAP Sequence, for Participants Exposed to Pro-Old and Pro-Young Exemplars, on Both Day 1 and Day 2

		Consistent						Inconsistent					
		Test 1		Test 2		Test 3		Test 1		Test 2		Test 3	
		M	SE	M	SE	M	SE	M	SE	M	SE	M	SE
Day 1													
Pro-Old	Con 1st	1,935	179	1,897	150	1,897	137	1,994	160	1,814	187	1,898	158
	Incon 1st	1,925	242	1,799	198	1,799	135	1,757	102	1,698	105	1,626	96
Pro-Young	Con 1st	2,004	134	1,843	130	1,864	103	2,135	174	2,050	146	2,092	131
	Incon 1st	1,915	94	1,878	68	1,841	56	2,296	99	1,946	118	1,986	118
Day 2													
Pro-Old	Con 1st	1,794	152	1,665	155	1,644	123	1,737	194	1,681	212	1,798	148
	Incon 1st	1,702	167	1,896	164	1,654	106	1,571	115	1,600	93	1,498	56
Pro-Young	Con 1st	1,708	94	1,663	103	1,621	69	1,804	120	1,838	123	1,882	129
	Incon 1st	1,668	83	1,819	113	1,678	81	1,822	150	1,899	162	1,964	173

viduals. The exemplar task was a tripartite exercise in which each participant was given (a) a practice booklet, (b) a test booklet, and (c) an age-test booklet, to be completed in that order. If fewer than 16 of the 20 participants failed to identify the correct profile and age (young or old) across the test and age booklets, participation in the experiment was terminated. Seven participants were removed on this basis.

Implicit measure. Participants completed an IRAP identical to that employed in Experiment 1, and then continued to Phase 3.

Explicit attitude measures. Participants completed the same semantic differentials and feeling thermometers that were employed in Experiment 1.

Implicit and explicit measures were administered 24 h later. On Day 2, each participant completed the same IRAP and self-report measures again, without any exposure to, or reminders of, the exemplars seen during the previous day.

Results and Discussion

Implicit Measure

The IRAP latency data were transformed using the same algorithm employed in Experiment 1. The overall mean latencies for each of the six IRAP test blocks for pro-young and pro-old groups, for Days 1 and 2, are presented in Table 3. Exposure to pro-young exemplars produced shorter latencies for consistent relative to inconsistent trials, both immediately after exemplar exposure and 24 h later. However, pro-old exemplars reversed this effect for both IRAP exposures.

IRAP administered immediately after exemplar exposure. Preliminary statistical analyses were conducted on the combined data set from the pro-young and pro-old groups to test for a possible order effect and to determine if it interacted with test sequence or IRAP condition. A $2 \times 2 \times 3$ mixed repeated measures ANOVA was applied to the adjusted latencies, with IRAP conditions (consistent vs. inconsistent) and test sequence (Tests 1, 2, and 3) as repeated measures and order (consistent-relations-first vs. inconsistent-relations-first) as the between-participants variable. The main and interaction effects for order were each nonsignificant ($ps > .49$) and thus order was not included in subsequent analyses. The main effect for test sequence proved to be significant, $F(2, 22) = 8.404$, $p = .0008$, $\eta_p^2 = .7$, but it did not interact with any of the other variables (all $ps > .42$). Given the significant main effect, test sequence was incorporated into subsequent analyses.

Following the preliminary analyses, a $2 \times 2 \times 3$ mixed repeated measures ANOVA was conducted with IRAP conditions (consistent vs. inconsistent) and test sequence (Tests 1, 2, and 3) as repeated measures and exemplars (pro-old vs. pro-young) as the between-participants variable. The main effect for test sequence was significant, $F(1, 22) = 8.276$, $p = .009$, $\eta_p^2 = .79$, as was the critical interaction effect for IRAP condition by exemplars, $F(1, 22) = 5.823$, $p = .02$, $\eta_p^2 = .001$. All remaining effects were nonsignificant (all $ps > .2$). The differential effects of exemplar training are presented in Figure 3 (panel A). The graph shows that the pro-young group produced longer latencies for inconsistent, relative to consistent, trials, but this pattern was reversed for the pro-old group.

Given the significant interaction, two separate 2×3 ANOVAs were conducted (IRAP condition test sequence), one for pro-young and one for pro-old. The pro-young ANOVA yielded a significant main effect for IRAP condition, $F(1, 11) = 10.124$, $p = .009$, $\eta_p^2 = .48$, and for test sequence, $F(2,$

11) = 6.253, p = .007, η_p^2 = .36. Critically, however, there was no interaction effect (p = .37), suggesting that the sequence effect was due simply to a practice effect across both consistent and inconsistent trials. The ANOVA for the pro-old group yielded no significant effects (ps > .1). In other words, the pro-young exemplars appeared to produce an IRAP performance similar to that produced in Experiment 1 with no exemplar training, whereas the pro-old exemplars removed the anti-old IRAP effect.

Additional analyses. Separate analyses were conducted for pro-young and pro-old groups to explore the effects of sample ("old people" vs. "young people") and target type (positive vs. negative), and to determine if there were any interaction effects among these variables and the critical IRAP condition
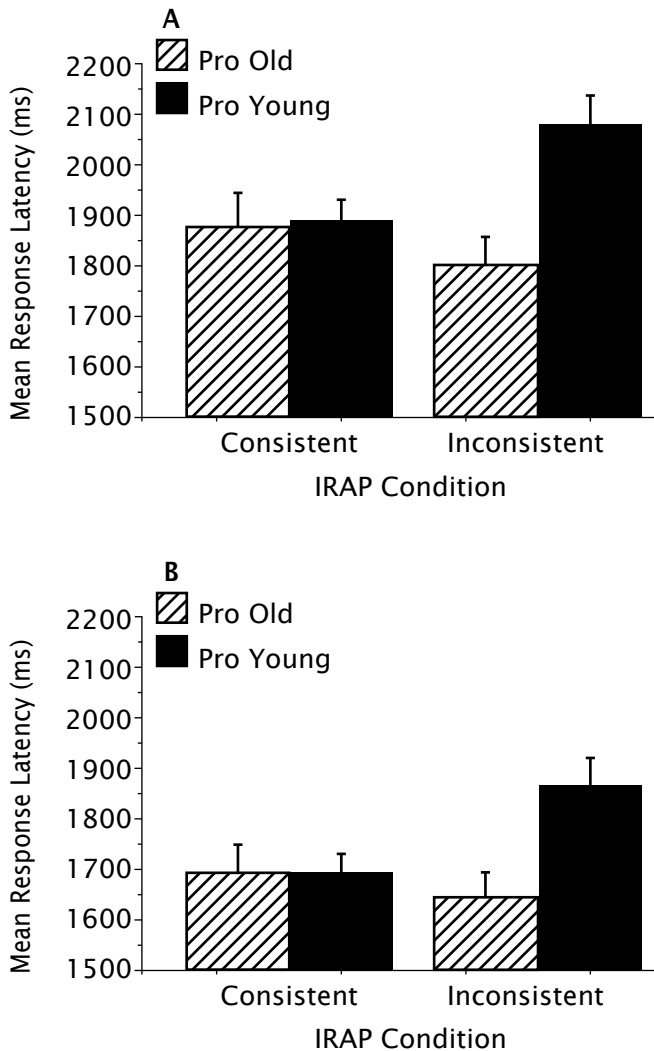

Figure 3. Response latencies following pro-old versus pro-young exemplars on the IRAP immediately after exemplar exposure (A) and 24 h later (B). Bars indicate standard errors.

(see Table 4). The pro-young exemplar group produced shorter response latencies for consistent relative to inconsistent trials across all combinations of samples and targets. The pro-old group also produced shorter latencies on consistent relative to inconsistent trials for "young people," but the reverse pattern was observed for the "old people" sample for both types of target.

Two separate $2 \times 2 \times 2$ ANOVAs were conducted, one for pro-young and one for pro-old, with sample, target type, and IRAP condition as repeated measures variables. The pro-young ANOVA yielded significant main effects for sample, $F(1, 11) = 12.919$, $p = .004$, $\eta_p^2 = .39$; target type, $F(1, 11) = 6.466$, $p = .03$, $\eta_p^2 = .37$; and IRAP condition, $F(1, 11) = 9.977$, $p = .009$, $\eta_p^2 = .48$. The effect for sample x IRAP condition was significant, $F(1, 11) = 5.542$, $p = .04$, $\eta_p^2 = .34$, with the interaction for sample $\times$ target type marginally so, $F(1, 11) = 4.636$, $p = .054$, $\eta_p^2 = .30$. The remaining two interaction effects were nonsignificant ($ps > .35$). The critical interaction effect between sample and IRAP condition is presented in Figure 4 (top-left panel). The data indicate that the difference between consistent and inconsistent trials was greater for "young people" than for "old people" samples. More informally, the participants found it easier to respond positively to young people than to respond negatively to old people.

Table 4
Overall Mean Response Latencies (in Milliseconds) Calculated Across All IRAP Trials, Divided in Terms of Sample, Target Type, Exemplar, and IRAP Condition on Both Day 1 and Day 2

		Young People				Old People			
		Consistent		Inconsistent		Consistent		Inconsistent	
		M	SE	M	SE	M	SE	M	SE
Day 1									
Pro-Old	Positive	1,742	132	1,772	101	1,921	112	1,733	96
	Negative	1,819	111	1,857	102	1,873	111	1,752	114
Pro-Young	Positive	1,746	68	1,955	68	1,981	72	2,098	108
	Negative	1,880	69	2,171	100	1,992	65	2,132	96
Day 2									
Pro-Old	Positive	1,573	91	1,644	109	1,713	91	1,628	99
	Negative	1,688	106	1,722	91	1,713	91	1,627	99
Pro-Young	Positive	1,582	48	1,829	93	1,741	60	1,931	103
	Negative	1,664	80	1,876	98	1,714	61	1,958	97

In contrast to the pro-young ANOVA, the pro-old ANOVA yielded only one significant effect: an interaction for sample $\times$ IRAP condition, $F(1, 11) = 20.63$, $p = .0008$, $\eta_p^2 = .65$. The effects of this interaction are presented in Figure 4 (top-right panel). The data indicate that the difference between consistent and inconsistent trials was relatively small for the "young people" samples, although it was in a pro-young direction. Critically, however, the consistent–inconsistent difference was relatively large and in the opposite direction for the "old people" samples. More informally, the participants did not show a strong preference for young people, but did show a relatively strong preference for old.

In summary, therefore, the pro-young exemplars produced a strong pro-young effect with a somewhat weaker anti-old effect. In contrast, the pro-old exemplars produced a very weak pro-young effect, but a relatively strong pro-old effect. To compare exemplar training directly, difference scores were calculated for each participant by subtracting mean latencies for consistent trials from inconsistent trials for "young people" samples, and similarly for "old people" samples. An independent t test was used to compare the difference scores for pro-young versus pro-old exemplar conditions for each sample (setting alpha at .025 to control for Type 1 error when conducting two t tests). The "Old-People" sample yielded a significant effect ($t = -2.6$, $p = .02$), with the "Young-People" sample approaching significance ($t = -1.9$, $p = .07$). These analyses support the conclusion that the exemplar training had a stronger impact on the old- than on the young-people IRAP effect.
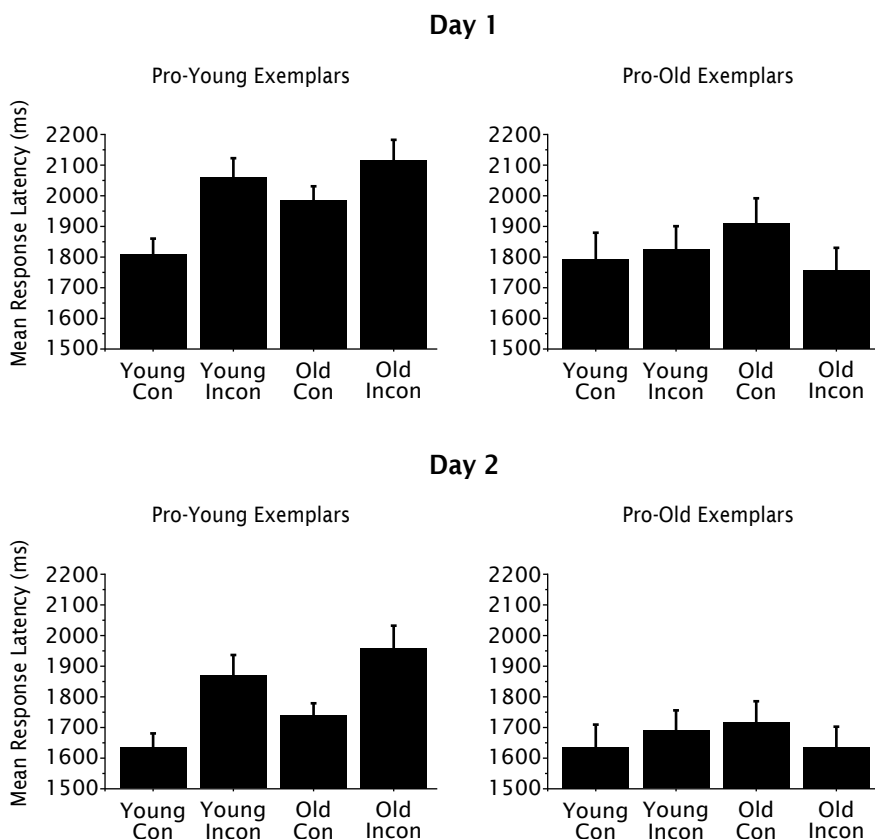

Figure 4. Response latencies following pro-old versus pro-young exemplars for "young people" and "old people" samples on the IRAP immediately after exemplar exposure (A) and 24 h later (B). Bars indicate standard errors. Con = consistent; Incon = inconsistent.

IRAP administered 24 h after exemplar exposure. The data from the Day 2 IRAP for both groups combined were subjected to the same preliminary statistical analyses as were applied to the Day 1 data (to test for an order effect and for interactions with test sequence or IRAP condition). The ANOVA yielded no such main or interaction effects, and thus order was not included

in subsequent analyses. Although the main effect for test sequence was not significant, it was incorporated into subsequent analyses to maintain analytic consistency with Day 1, and, furthermore, the test sequence × IRAP condition interaction did approach significance ($p = .07$; all other $ps > .2$).

A $2 \times 2 \times 3$ mixed repeated measures ANOVA, with IRAP condition and test sequence as repeated measures and exemplars as the between-participants variable, indicated no significant main effects (all $ps > .2191$). Similar to Day 1, however, the critical interaction effect for IRAP condition by exemplars was significant, $F(1, 21) = 5.533$, $p = .03$, $\eta_p^2 = .21$. All remaining interaction effects were nonsignificant (all $ps > .09$). The graphs presented in Figure 3 (panel B) illustrate that the general pattern of results obtained in Day 1 were also observed in Day 2, with an overall reduction in response latencies, probably due to a simple practice effect, across the 2 days. Two separate 2×3 ANOVAs were conducted (IRAP condition × test sequence), one for pro-young and one for pro-old. Similar to Day 1, the pro-young ANOVA yielded a significant effect for IRAP condition, $F(1, 11) = 8.222$, $p = .01$, $\eta_p^2 = .43$, with no interaction effect ($p = .16$). Unlike Day 1, there was no effect for test sequence ($p = .36$), supporting the conclusion that the Day 1 effect was due to practice. The ANOVA for the pro-old group yielded no significant effects, as was the case for Day 1 (all $ps > .39$).

Additional analyses. The same additional analyses that were conducted for the Day 1 data were conducted for Day 2 (see Table 4), and the same pattern of results emerged. Specifically, the pro-young group produced pro-young IRAP effects across both combinations of samples and targets, whereas this effect was observed only for the "young people" sample, and was reversed for the "old people" sample, for the pro-old group.

Similar to Day 1, the pro-young $2 \times 2 \times 2$ ANOVA yielded a significant effect for the critical IRAP condition, $F(1, 11) = 10.892$, $p = .007$, $\eta_p^2 = .50$. Unlike Day 1, however, there was no interaction between sample and IRAP condition ($p = .87$). A comparison of the upper and lower panels of Figure 4 for the pro-young exemplars suggests that the interaction effect disappeared in Day 2 because the difference between consistent and inconsistent trials for the "old people" sample increased in size from Day 1 to Day 2.

The ANOVA for the pro-old data yielded three significant effects: one for target type, $F(1, 10) = 7.780$, $p = .02$, $\eta_p^2 = .45$; an interaction for sample × target type, $F(1, 10) = 7.780$, $p = .02$, $\eta_p^2 = .45$; and an interaction for sample × IRAP condition, $F(1, 10) = 8.433$, $p = .016$, $\eta_p^2 = .46$. Neither of the former two effects was obtained for Day 1, but they remain relatively unimportant because they did not interact with the critical IRAP condition ($ps > .73$). Consistent with Day 1, the principal finding was the latter interaction effect.

In summary, therefore, 24 h after exemplar exposure, the pro-young exemplars continued to produce a strong bias toward young over old people, with some suggestion that the bias against old people actually increased. The weak pro-young effect observed for the pro-old exemplar group was maintained across the 2 days, but the relatively strong pro-old bias seen in Day 1 was reduced somewhat in Day 2. To compare exemplar training directly for Day 2, difference scores were calculated as for Day 1. Once again, an independent t test revealed that the exemplars significantly affected the difference scores for the "old people" sample ($t = -2.528$, $p = .02$) but not for the "young people" sample ($t = -1.680$, $p = .1$). Overall, therefore, the general pattern of results observed in Day 1 was observed in Day 2 for both groups.

Explicit Measures

Explicit measures administered immediately after the IRAP. The data from the explicit measures were subjected to the same statistical analyses as the data from Experiment 1. Similar to the previous experiment, participants expressed a greater liking for young compared with old people on semantic differential scales ($Ms = 1.493$ and -0.180; $d = .98$). A 2×2 mixed repeated measures ANOVA with explicit rating (old vs. young) as the repeated measure and exemplar (pro-old vs. pro-young) as the between-participants variable indicated that the preference for young over old was significant, $F(1, 22) = 21.021$, $p = .0001$, $\eta_p^2 = .48$. The main effect for exemplar was nonsignificant, as was the interaction effect ($ps > .23$). Unlike Experiment 1, participants reported slightly less favorable attitudes toward young compared with old people on the feeling thermometers ($Ms = 66.67°$ and $70.83°$; $d = -.42$), but the difference between these ratings was not significant ($p = .1517$). Once again, the main effect for exemplar was nonsignificant, as was the interaction effect ($ps > .2289$).

Explicit measures administered 24 h after exemplar exposure. The pattern of explicit data remained stable 24 h after encountering the exemplars on both semantic differential scales and feeling thermometers. Similar to Day 1, participants again reported more favorable attitudes toward young compared with old people on the semantic differentials ($Ms = 1.492$ and $.334$; $d = .70$), and this difference was significant, $F(1, 21) = 10.971$, $p = .003$, $\eta_p^2 = .35$. Consistent with Day 1, no other semantic differential effects were significant ($ps > .47$). Participants continued to report slightly more favorable attitudes toward old compared with young people on feeling thermometers on Day 2 ($Ms = 70.87°$ and $70.435°$; $d = -.03$), but again this difference was not significant ($p = .86$). Consistent with Day 1, no other effects were significant for the thermometers ($ps > .35$). Overall, the general pattern of results observed in Day 1 was observed in Day 2.

Relationship Between Implicit and Explicit Measures

Difference scores for the IRAP, semantic differential, and feeling thermometer measures were calculated as per Experiment 1. No significant correlations were found between the explicit and implicit measures for either Day 1 or Day 2 (Day 1: IRAP and feeling thermometers, $r = -0.086$, $p = .68$; IRAP and semantic differentials, $r = -0.006$, $p = .98$: Day 2: IRAP and feeling thermometers, $r = -0.266$, $p = .21$; IRAP and semantic differentials, $r = -.110$, $p = .61$). Similarly, no significant correlations existed between the two explicit measures for ratings of either old or young people for either of the two days (Day 1: old people, $r = -0.104$, $p = .62$; young people, $r = 0.263$, $p = .21$: Day 2: old people, $r = .315$, $p = .14$; young people, $r = .300$, $p = .16$).

Test–Retest Reliability

Correlations revealed that participants' responses on implicit measures and on feeling thermometers remained remarkably stable across the two experimental sessions: IRAP, $r = .49$, $p = .0215$; feeling thermometer, $r = .688$, $p = .0012$ (young people ratings), $r = .804$, $p = .0002$ (old people ratings). In contrast, the semantic differential measure did not remain as stable across sessions: $r = .234$, $p = .27$ (young people ratings), $r = .266$, $p = .21$ (old people ratings).

Summary

Pro-young exemplar exposure produced an IRAP performance similar to that produced in Experiment 1 with no exemplar training. Although there was an overall reduction in response latencies on Day 2, suggestive of a practice effect, the pro-young group produced longer latencies for inconsistent, relative to consistent, trials on both days. In contrast, the pro-old exemplar condition produced the opposite pattern—a pro-old IRAP effect across both exposures. Additional analyses revealed that on Day 1 and Day 2, the pro-young exemplar condition produced a strong pro-young effect with a somewhat weaker anti-old effect, with Day 2 showing some evidence of an increased bias against old people. In contrast, the pro-old exemplar condition produced a very weak pro-young effect but a relatively strong pro-old effect, with some reduction in the latter in Day 2. The explicit measures remained stable across the 2 days, with a significant pro-young effect for the semantic differentials regardless of exemplar condition; the feeling thermometer effects were nonsignificant. The correlations between implicit and explicit measures were all weak and nonsignificant. The test–retest reliability was positive and significant for the IRAP and feeling thermometers, but non–significant for the semantic differentials.

General Discussion

The results from the IRAP in Experiment 1 revealed a bias toward young people, and thus it performed in a way that was broadly similar to the IAT, in that a pro-young bias was observed in the absence of any programmed exemplar training (see Jelenec & Steffens, 2002; Nosek, Banaji, & Greenwald, 2002; Rothermund & Wentura, 2001). In the second experiment, the pro-young exemplars produced an IRAP performance that was broadly similar, across both days, to that produced in Experiment 1 with no exemplar training (i.e., a pro-young bias). In contrast, the pro-old exemplar condition reversed the overall IRAP effect, with longer response latencies for inconsistent relative to consistent trials, although the difference was nonsignificant. Once again, therefore, the IRAP performed in a way that is broadly consistent with the IAT study reported by Dasgupta and Greenwald (2001).

The correlations between implicit and explicit measures for Experiment 2 of the current study were all weak and nonsignificant. This finding is consistent with the Dasgupta and Greenwald (2001) study. Furthermore, the test–retest reliability was positive and significant for the IRAP. A similar result was obtained for the feeling thermometers, but not for the semantic differentials (i.e., the positive correlation for the latter was nonsignificant). A direct comparison of the IRAP test–retest reliability with that of the IAT is not possible because the age-related experiment reported by Dasgupta and Greenwald involved a single exposure to the IAT.

Consistent with Dasgupta and Greenwald (2001), neither the pro-young nor the pro-old exemplars in the current study appeared to significantly affect the feeling thermometer measure for either exposure. Interestingly, Dasgupta and Greenwald reported an unexpected effect for pro-old exemplars on their semantic differential scales, in that this condition produced an explicit preference for old over young; the converse effect was not observed for the pro-young condition, which produced no preference. In the current study,

however, a significant preference for young over old was observed in *both* exemplar conditions. A likely explanation for this difference across the two studies is the fact that the terms employed by Dasgupta and Greenwald for the semantic differentials were relatively extreme: *ugly–beautiful, bad–good, unpleasant–pleasant, dishonest–honest,* and *awful–nice.* In contrast, the following milder terms were employed in the current study: *creative–stagnant, happy–sad, brilliant–slow, energetic–tired, enthusiastic–weary*, and *productive–finished.* The latter terms, therefore, may have been less prone to social desirability effects. For example, reporting that old people are tired or weary (as opposed to ugly or bad) might not be seen as particularly prejudiced or subject to social opprobrium. Consequently, the current pro-old exemplars may have had a minimal impact on the semantic differentials because participants did not consider their negative responses to old people as particularly negative. It is also important to note that in the current study, the same target terms were used for both the explicit and implicit measures, whereas Dasgupta and Greenwald employed different terms across the two sets of measures. Consequently, the explicit–implicit divergence observed in the current research provides a particularly powerful example of implicit cognition.

The current results indicate that attitudes measured using a latency-based procedure may diverge from the results of explicit measures and that the former appear to be more malleable than the latter. Furthermore, the IRAP indicates that the effect of pro-old exemplar training differentially affected implicit attitudes to *young* and *old* by weakening the pro-young and reversing the anti-old bias.[4] The fact that the pro-old exemplar condition produced these two separate effects indicates that the IRAP, in contrast to the IAT, does indeed permit the assessment of attitudes toward individual concepts. More specifically, the reversal in the overall IRAP effect for the pro-old condition in Experiment 2 indicated not the establishment of an anti-young bias, but a reversal in the anti-old effect. As explained in the Introduction, the IAT cannot provide this level of analytic detail for individual concepts, and thus it is not possible to determine if the pro-old exemplars had a similar effect in the Dasgupta and Greenwald (2001) study. Indeed, the same interpretive problem could apply to any study of malleability that employs the IAT as an implicit measure, because the effects on the individual categories are inseparable (cf. Blair, 2002). At the very least, therefore, the current findings indicate that the IRAP could provide a possibly useful alternative to the IAT when a fine-grained analysis of implicit cognition is required.

The Relational Elaboration and Coherence Model: A Relational Frame Theory Explanation for the Divergence Between Implicit and Explicit Measures

In the Introduction, we predicted that the IRAP effect, similar to the IAT, would prove sensitive to the pro-old exemplar exposure, and this appeared

---

4 The differential impact observed in Experiment 2 may be explained as follows. The participants were all relatively young university students, and recent IRAP research indicates that this group has high levels of implicit self-esteem (Vahey, et al., in press). Furthermore, in-group favoritism and self-esteem are positively correlated (Farnham, Greenwald, & Banaji, 1999). Perhaps, therefore, the anti-young exemplars, which targeted the participants' in-group, had a lesser impact than the pro-old exemplars, which targeted an out-group (i.e., the exemplar effect on the in-group was attenuated by high self-esteem).

to be the case. Interestingly, however, the explicit measures appeared to be unaffected. To explain this divergence in the two measures from the perspective of RFT, we offer below what we call the Relational Elaboration and Coherence (REC) model (Barnes-Holmes, Barnes-Holmes, Stewart, & Boles, in press).

The REC model assumes that specific IRAP trials may produce an immediate and relatively brief relational response before the participant actually presses a response key. By definition, the most probable immediate response will be emitted first most often, and thus any IRAP trial that requires a key press that coordinates with that immediate response will be emitted relatively quickly; if, however, an IRAP trial requires a key press that opposes the immediate relational response, then it may be emitted less quickly. Critically, the REC model assumes that the probability of the initial response on the IRAP will often be determined by the verbal and nonverbal history of the participant and current contextual variables. Thus, participants in Experiment 1 produced IRAP performances that appeared to reflect exposure to pre-experimental verbal and nonverbal behavioral contingencies (i.e., responding more quickly to *Young-Positive* and *Old-Negative* stimulus relations, but more slowly to their reversed counterparts). When participants were first exposed to repeated examples of negative young people and positive old people, however, this appeared to establish the IRAP, and the experimental setting more generally, as contexts that increased the probability of atypical immediate relational responding.

In attempting to explain why the exemplars did not have an impact on the explicit measures, the REC model assumes that responses to these measures likely reflected relatively elaborate and coherent relational responding. In other words, when asked to express an attitude or belief on a particular issue, it is likely that a person will produce a relational response that coheres with one or more other relational responses in his or her behavioral repertoire (see Barnes-Holmes, Hayes, & Dymond, 2001). Imagine, for example, that a participant indicated that "old people are tired" on a semantic differential. This simple relational response would likely cohere with other relevant relational networks, such as "My grandma falls asleep in her chair when we visit," and "The old man next door can't walk far without taking a rest." The critical point here is that explicit measures typically are not completed under high time pressure, and thus participants have sufficient time to engage in the extended relational responding that is needed to produce a response that coheres with one or more other relational responses. When exposed to the IRAP, however, the impact of a participant's elaborated relational responding would be absent or much reduced because there is insufficient time, on a trial-by-trial basis, to engage in the additional and sometimes complex relational activity that serves to generate a relationally coherent response.

In summary, therefore, the REC model assumes that the IRAP effect is largely driven by immediate and relatively brief relational responses, whereas explicit measures reflect extended and coherent relational networks. Or, more informally, the IRAP captures spontaneous and automatic evaluations, whereas explicit measures capture more carefully considered reactions. The core of the REC model explanation for the divergence between the two measures in the current study rests on the following two assumptions. First, immediate or automatic evaluative responses may or may not cohere with subsequent relational responding; when they cohere, implicit and explicit

measures will typically converge, but when they do not, the measures will typically diverge. In other words, it is assumed that participants typically "reject" their immediate and brief relational responses (or automatic evaluations) if they do not cohere with their more elaborate and extended relational responding. Second, the REC model assumes that simple repeated stimulus exposures, such as the exemplar training employed in the current study, would impact largely on immediate relational responding but less so on extended and coherent relational networks. Thus, the pro-old exemplars in the current study may have served to evoke old-positive immediate relational responses on the IRAP, but these were subsequently "rejected" when the explicit measures were completed because they did not cohere with elaborated relational responding. More informally, when asked to rate old people in terms of tired-versus-energetic, some of the old-positive exemplars may well have come to mind, but these were rejected as the basis for a rating because they were deemed atypical (e.g., "Einstein was brilliant, but most old people are still less energetic than most young people I know").

Admittedly, the REC model explanation for the divergence between explicit and implicit measures is highly speculative at the current time.[5] Nevertheless, it does appear to provide a possibly useful account that may be subjected to experimental analysis in future studies. For example, the REC model would predict that the divergence between implicit and explicit "socially sensitive" attitudes should increase with greater time pressure on the IRAP (because participants have less time to engage in elaborated relational responding). In any case, the current findings have shown that implicit attitudes, measured with an IRAP, appear to be more sensitive than explicit attitudes to relevant exemplars. Only future research will determine if the REC model, or an alternative account, will fully explain these and related findings.

References

ALLPORT, G. W. (1954). *The nature of prejudice*. Reading, MA: Addison-Wesley.

BANSE, R., SEISE, J., & ZERBES, N. (2001). Implicit attitudes towards homosexuality: Reliability, validity, and controllability of the IAT. *Zeitschrift fur Experimentelle Psychologie, 48*, 145–160.

BARGH, J. A. (1999). The cognitive monster: The case against the controllability of automatic stereotype effects. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 361–382). New York: Guilford.

BARNES-HOLMES, D., BARNES-HOLMES, Y., POWER, P., HAYDEN, E., MILNE, R., & STEWART, I. (2006). Do you really know what you believe? Developing the Implicit Relational Assessment Procedure (IRAP) as a direct measure of implicit beliefs. *The Irish Psychologist, 32*, 169–177.

---

5 The REC Model bears some similarity to the Associative-Propositional Evaluation Model (APE model; e.g., Gawronski & Bodenhausen, 2007). Unlike the APE model, however, the REC Model does not appeal to dual processes (associative and propositional). Rather, the REC model appeals to the single process of arbitrarily applicable relational responding, as defined by RFT. Thus, the divergence between implicit and explicit attitudes is explained not by the distinction between associative and propositional processes but by the extent to which relational responses are elaborated and cohere. Furthermore, the REC model predicts that implicit attitude effects are not restricted to simple associations, but may emerge based on a variety of stimulus relations (see Power, et al., in press, for supporting evidence).

BARNES-HOLMES, D., BARNES-HOLMES, Y., STEWART, I., & BOLES, S. (in press). A sketch of the Implicit Relational Assessment Procedure (IRAP) and the Relational Elaboration and Coherence (REC) model. *The Psychological Record.*

BARNES-HOLMES, D., HAYDEN, E., BARNES-HOLMES, Y., & STEWART, I. (2008). The Implicit Relational Assessment Procedure (IRAP) as a response-time and event-related-potentials methodology for testing natural verbal relations: A preliminary study. *The Psychological Record, 58*, 497–516.

BARNES-HOLMES, D., HAYES, S. C., & DYMOND, S. (2001). Self and self-directed rules. In S. C. Hayes, D. Barnes-Holmes, & B. Roche (Eds.), *Relational Frame Theory: A post-Skinnerian account of human language and cognition* (pp. 119–140). New York: Plenum.

BARNES-HOLMES, D., MURTAGH, L., BARNES-HOLMES, Y., & STEWART, I. (in press). Using the Implicit Association Test and the Implicit Relational Assessment Procedure to measure attitudes towards meat and vegetables in vegetarians and meat-eaters. *The Psychological Record.*

BARNES-HOLMES, D., WALDRON, D., BARNES-HOLMES, Y., & STEWART, I. (2009). Testing the validity of the Implicit Relational Assessment Procedure (IRAP) and the Implicit Association Test (IAT): Measuring attitudes towards Dublin and country life in Ireland. *The Psychological Record, 59, 389–406.*

BLAIR, I. V. (2002). The malleability of automatic stereotypes and prejudice. *Personality and Social Psychology Review, 6*, 242–261.

BLAIR, I. V., & BANAJI, M. R. (1996). Automatic and controlled processes in stereotype priming. *Journal of Personality and Social Psychology, 70*, 1142–1163.

BOSSON, J. K., SWANN, W. B., & PENNEBAKER, J. W. (2000). Stalking the perfect measure of implicit self-esteem: The blind men and the elephant revisited? *Journal of Personality and Social Psychology, 79*, 631–643.

BOYSEN, G. A., VOGEL, D. L., & MADON, S. (2006). A public versus private administration of the Implicit Association Test. *European Journal of Social Psychology, 36*, 845–856.

COHEN, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

DASGUPTA, N., & GREENWALD, A. G. (2001). On the malleability of automatic attitudes: Combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology, 81*, 800–814.

DE HOUWER, J. (2002). The Implicit Association Test as a tool for studying dysfunctional associations in psychopathology: Strengths and limitations. *Journal of Behavior Therapy and Experimental Psychiatry, 33*, 115–133.

DE HOUWER, J. (2003). The Extrinsic Affective Simon task. *Experimental Psychology, 50*, 77–85.

DE JONG, P. (2002). Implicit self-esteem and social anxiety: Differential self-positivity effects in high and low anxious individuals. *Behaviour Research and Therapy, 40*, 501–508.

DEVINE, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology, 56*, 5–18.

DEVINE, P. G., MONTEITH, M. J., ZUWERINK, J. R., & ELLIOT, A. J. (1991). Prejudice with and without compunction. *Journal of Personality and Social Psychology, 60*, 817–830.

FARNHAM, S. D., GREENWALD, A. G., & BANAJI, M. (1999). Implicit self-esteem. In D. Abrams & M. Hogg (Eds.), *Social identity and social cognition* (pp. 230–248). Oxford, UK: Blackwell.

FAZIO, R. H., SANBONMATSU, D. M., POWELL, M. C., & KARDES, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology, 50*, 229–238.

GAWRONSKI, B., & BODENHAUSEN, G. V. (2007). Unraveling the processes underlying evaluation: Attitudes from the perspective of the APE model. *Social Cognition, 25*, 687–717.

GREENWALD, A. G., & BANAJI, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102*, 4–27.

GREENWALD, A. G., BANAJI, M. R., RUDMAN, L. A., FARNHAM, S. D., NOSEK, B. A., & MELLOTT, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review, 109*, 3–25.

GREENWALD, A. G., MCGHEE, D. E., & SCHWARTZ, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology, 74*, 1464–1480.

GREENWALD, A. G., NOSEK, B. A., & BANAJI, M. R. (2003). Understanding and 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., BANAJI, M. R., & KLAUER, K. C. (2005). Validity of the salience asymmetry interpretation of the Implicit Association Test: Comment on Rothermund and Wentura (2004). *Journal of Experimental Psychology: General, 134*, 420–425.

HAYES, S. C., BARNES-HOLMES, D., & ROCHE, B. (2001). *Relational Frame Theory: A post-Skinnerian account of human language and cognition.* New York: Plenum.

JELENEC, P., & STEFFENS, M. C. (2002). Implicit attitudes towards elderly women and men. *Current Research in Social Psychology, 7*, 275–293.

KAWAKAMI, K., DOVIDIO, J. F., MOLL, J., HERMSEN, S., & RUSSIN, A. (2000). Just say no (to stereotyping): Effect of training in the negation of stereotypic associations on stereotype activation. *Journal of Personality and Social Psychology, 78*, 871–888.

LEPORE, L., & BROWN, R. (1997). Category and stereotype activation: Is prejudice inevitable? *Journal of Personality and Social Psychology, 72*, 275–287.

LOWERY, B., HARDIN, C. D., & SINCLAIR, S. (2001). Social tuning effects on automatic racial prejudice. *Journal of Personality and Social Psychology, 81*, 842-855.

MCKENNA, I. M., BARNES-HOLMES, D., BARNES-HOLMES, Y., & STEWART, I. (2007). Testing the fake-ability of the Implicit Relational Assessment Procedure (IRAP): The first study. *International Journal of Psychology and Psychological Therapy, 7*, 123-138.

MOSKOWITZ, G. B., SALOMON, A. R., & TAYLOR, C. M. (2000). Preconsciously controlling stereotyping: Implicitly activated egalitarian goals prevent the activation of stereotypes. *Social Cognition, 18*, 151-177.

MYRDAL, G. (1944). *An American dilemma: The Negro problem and modern democracy.* New York: Harper.

NOSEK, B., & BANAJI, M. R. (2001). The go/no-go association task. *Social Cognition, 19*, 625-666.

NOSEK, B. A., BANAJI, M., & GREENWALD, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics, 6*, 101-115.

NOSEK, B., GREENWALD, A. G., & BANAJI, M. R. (2004). Understanding and using the Implicit Association Test: II. Method variables and construct validity. *Personality and Social Psychology Bulletin, 31*, 166-180.

POWER, P. M., BARNES-HOLMES, D., BARNES-HOLMES, Y., & STEWART, I. (in press). The Implicit Relational Assessment Procedure (IRAP) as a measure of implicit relative preferences: A first study. *The Psychological Record.*

PRATTO, F., & JOHN, O. P. (1991). Automatic vigilance: The attention grabbing power of negative social information. *Journal of Personality and Social Psychology, 61*, 380-391.

ROTHERMUND, K., & WENTURA, D. (2001). Figure-ground asymmetries in the Implicit Association Test (IAT). *Zeitschrift fuer Experimentelle Psychologie, 48*, 94-106.

RUDMAN, L. A., & GLICK, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues, 57*, 743-762.

RUDMAN, L. A., GREENWALD, A. G., MELLOTT, D. S., & SCHWARTZ, J. L. K. (1999). Measuring the automatic components of prejudice: Flexibility and generality of the Implicit Association Test. *Social Cognition, 17*, 437-465.

VAHEY, N. A., BARNES-HOLMES, D., BARNES-HOLMES, Y., & STEWART, I. (2009). A first test of the Implicit Relational Assessment Procedure (IRAP) as a measure of self-esteem: Irish prisoner groups and university students. *The Psychological Record, 59, 371-388.*

Appendix

Old and Young Exemplars Used in Experiment 2

Name	True Description	False Description
	Admired Old Exemplars	
Ronnie Drew	Irish singer and folk musician who co-founded the 'Dubliners'.	Former paediatrician famous for his books on child rearing.
Albert Einstein	World-famous physicist who won the Nobel Prize.	World-famous psychic helped FBI solve cases.
Mother Teresa	Was a true champion of the poor and destitute.	Famous novelist and poet.
Mary Robinson	Was the first female President of Ireland and the United Nations High Commissioner for Human Rights.	A popular chat show host on RTE Radio 1, who co-presented with Derek Davis.
Jack Charlton	Managed the Republic of Ireland Football Team and took them to the World Cup in 1990.	Performance artist and poet. Designed the 'Spike' in Dublin's O'Connell Street.
Maureen Potter	Acclaimed Irish comedienne and actress.	Horticulturalist famous for her programme 'Gardener's World'.
Frank Sinatra	One of the most influential singers of the 20th century.	Irish actor who played St. Francis of Assisi in a Gate theatre production.
Angela Lansbury	Famous Hollywood actress who starred in "Miss Marple," "Murder on the Orient Express" and "Murder She Wrote."	Mother of Robbie Williams who is famous for her charity work in the UK.
Maureen O'Hara	Irish American film actress awarded 'the Lifetime Achievement Award' by the Irish Film & Television Academy in 2004.	Founder of the Irish Country Woman's Association and winner of the Calor Kosangas 'House Wife the Year' competition in 1988.
Gay Byrne	Presented the 'Late Late Show' — the world's longest running chat show.	Prize-winning journalist for the *Irish Times* newspaper.

Old and Young Exemplars Used in Experiment 2

Name	True Description	False Description
	Disliked Old Exemplars	
Fr. Brendan Smyth	Irish paedophile priest jailed after pleading guilty to 74 cases of child abuse.	Had multiple adulterous affairs with Irish housewives between 1976 and 1988.
Pol Pot	Cambodian leader who killed two million people through mass execution.	Former Viet-Cong Colonel tried for war crimes.
Harold Shipman	The world's most prolific serial killer.	Hit and run driver who mowed down a young mother and baby.
Charles Haughey	Ex-Taoiseach who corruptly used taxpayers' money for his own ends.	Paedophile farmer from Wicklow currently in hiding in South America.
Margaret Thatcher	Former UK Prime Minister whose policies caused widespread unemployment and dissent.	Wife of Romanian dictator found guilty of crimes against humanity.
Myra Hindley	Known as 'The Moors Murderess'. She received two life sentences for the rape, torture, and murder of 5 children.	Gun-runner for the IRA who was shot dead in Guilford in 1978.
Catherine Nevin	Found guilty of murdering her husband.	Mother in Galway who drowned both of her children.
Jane O'Brien	County Wexford woman who shot her nephew in order to get possession of a farm.	Imprisoned for murdering and dismembering one of her brothers in County Cork.
Imelda Marcos	Wife of former Philippine President, who stole between 5 and 10 billion dollars of taxpayers' money to fund her shopping sprees.	Infamous brothel owner who held 30 girls against their will in Cork's inner city.
John Gilligan	Irish criminal implicated in the murder of journalist Veronica Guerin.	Day-care worker accused of neglecting children at his day-care centre.

Old and Young Exemplars Used in Experiment 2

Name	True Description	False Description
	Admired Young Exemplars	
Reneé Zellweger	Academy Award-winning movie actress who famously starred in 'Bridget Jones's Diary' in 2001.	Winner of three gold medals for running in the 2004 Athens Olympics.
Jodie Foster	Famous actress.	Famous novelist.
Dermot O'Leary	Presenter of 'Big Brother's Little Brother'.	Captain of the British Lions tour to New Zealand in 2005.
Beyoncé Knowles	American R&B solo singer and lead singer of 'Destiny's Child' crowned 'Princess of Pop' in 2004.	Famous 'Bond Girl' and four-time Academy Award nominee for 'Best Actress'.
Damien Duff	Irish international mid-fielder.	Wexford hurling champion.
Robbie Williams	British Pop singer and former member of 'Take That'.	Irish international soccer goalkeeper.
Samantha Mumba	Famous Irish Pop singer.	Rose of Tralee 2004.
David Beckham	English footballer who plays for Real Madrid and captains the English National team.	Celebrity hair-stylist and owner of the Toni & Guy chain of hair salons.
Craig Doyle	Irish television presenter.	Irish male model.
Nadine Coyle	Northern Irish singer and member of pop group 'Girls Aloud'.	Professional Golfer who won the 'Ladies Irish Open' in 2002 and 2004.

Old and Young Exemplars Used in Experiment 2

Name	True Description	False Description
	Disliked Young Exemplars	
Timothy McVeigh	The 'Oklahoma City Bomber'. He killed 168 people and injured 500 after bombing the Oklahoma City federal building.	School caretaker who stole thousands from the school where he worked.
Pete Doherty	Irish ex-frontman of 'The Libertines' and drug-addicted boyfriend of Kate Moss.	Olympic skater accused of plotting to injure his competitor.
Louise Woodward	Nanny who was convicted of child murder.	Convicted for dealing drugs.
Malcolm McArthur	Double murderer from County Meath, currently serving the second longest ever life term in Irish prisons for brutally killing young victims.	Irish intensive care nurse sentenced to 17 years for murdering six patients whom he claimed begged him to help them die.
Marie Murray	Sentenced to life imprisonment in 1976 for shooting a policeman in an armed bank raid in Northern Ireland.	Famously kidnapped a dentist and having removed his fingers, sent them to his wife.
Rebecca Loos	Former personal assistant to David Beckham who allegedly had an adulterous relationship with him.	Former charity worker found guilty of embezzlement and now serving 6 years in prison.
Michelle De Bruin	Olympic medallist banned from future competition after a urine sample was found to be contaminated.	Murderer who killed fashion designer Gianni Versace.
Martin Nash	Confessed to the Grangegorman cannibal murders.	Irish drug baron serving 18 years in Mountjoy.
Maxine Carr	Convicted for perverting the course of justice and providing a false alibi for the Soham murderer.	Bombed the federal building in Oklahoma City, killing hundreds.
Jeffrey Dahmer	Serial killer who cannibalized his victims.	Bombed the World Trade Center in New York City.