ORIGINAL ARTICLE

Using a Modified Function Acquisition Speed Test (FAST) for Assessing Implicit Gender Stereotypes

Aoife Cartwright[1] · Bryan Roche[1] · Michelle Gogarty[1] · Anthony O'Reilly[2] ·
Ian Stewart[3]

Abstract The recently developed Function Acquisition Speed Test (FAST) represents an effort to assess the relative strength of stimulus relations by traditional behavior-analytic means (i.e., acquisition curves). The current study was the first application of the FAST to the assessment of natural, pre-experimentally established stimulus relations. Specifically, this experiment assessed the sensitivity of the FAST to pervasive gender stereotypes of men as stereotypically "masculine" (e.g., dominant or competitive) and women as stereotypically "feminine" (e.g., nurturing or gentle). Thirty participants completed a FAST procedure consisting of two testing blocks. In one block, functional response classes were established between classes of stimuli assumed to be stereotype-consistent (i.e., men-masculine and women-feminine), and in the other, between classes of stimuli assumed to be stereotype-inconsistent (i.e., men-feminine and women-masculine). Differences in the rate of class acquisition across the two blocks were quantified using cumulative record-type scoring procedures plotting correct responses as a function of time. Acquisition rates were significantly faster (i.e., displayed steeper learning curves) for the stereotype-consistent relative to the stereotype-inconsistent block. Corroborating stereotypes were observed on an Implicit Association Test containing identical stimuli.

Relational Frame Theory (RFT: Hayes, Barnes-Holmes, & Roche, 2001) has shown promise over the last 20 years in conceptualizing verbal behavior from a functional-analytic, operant perspective. It has also demonstrated utility in modeling a range of complex verbal behaviors in the laboratory, such as analogical reasoning, self-awareness, and learning via instruction (see Hughes & Barnes-Holmes, 2016; Roche & Dymond, 2013). As yet, however, only a limited number of methodologies have been developed that are capable of assessing naturalistic relational responding (e.g., Hussey, Barnes-Holmes, & Barnes-Holmes, 2015). Most notably, the social-cognitive Implicit Association Test (IAT: Greenwald, McGhee, & Schwartz, 1998) and the behavior-analytic Implicit Relational Assessment Procedure (IRAP: Barnes-Holmes, Barnes-Holmes, Stewart, & Boles, 2011) have both been used and interpreted from within a behavior-analytic research paradigm to measure verbal histories and patterns of relational responding across a range of clinically and socially relevant domains (e.g., Gavin, Roche, Ruiz, Hogan, & O'Reilly, 2012; Ridgeway, Roche, Gavin, & Ruiz, 2010; Vahey, Nicholson, & Barnes-Holmes, 2015).

To provide a brief overview, based on the seminal finding that learning histories may interfere with the ability to form novel stimulus relations that run counter to them (see Watt, Keenan, Barnes, & Cairns, 1991), both the IRAP and the IAT attempt to infer some aspect of an individual's history by comparing their performance on two opposing tasks. In one task, participants are required to form relations assumed to be consistent with those arising from the dominant verbal community (e.g., women as submissive and men as dominant),

✉ Aoife Cartwright
aoife.m.cartwright@gmail.com

[1] Maynooth University, Kildare, Ireland

[2] University of Essex, Galway, Ireland

[3] National University of Ireland, Galway, Ireland

and in the other, relations that are deemed to be inconsistent (e.g., men as submissive and women as dominant). Any differences in performance across the two tasks (e.g., in the rate of fluency) are taken as indicative of the differential impact of a learning history on one task relative to the other, and thus, as a metric of that history itself. Inferring content from behavioral performances in this way (rather than self-reports) is what leads to many researchers referring to these measures as "implicit".

Both the IAT and the IRAP have employed a relatively similar strategy to date, which is to compare differences in the speed of relational responding (or categorizing, within the IAT literature) across two contrasting test blocks. In brief, the IAT involves a series of stimulus categorization tasks, in which participants are trained to respond with left and right key-presses to individual stimuli presented on-screen. Participants are first required to categorize two target categories (e.g., men/women) followed by two attributes (e.g., dominant/submissive), before categorizing targets and attributes together simultaneously (e.g., men-dominant and women-submissive). In the critical test blocks, reinforcement was contingent upon responding first in a way that is assumed to be culturally consistent (e.g., men-dominant and women-submissive) and secondly, in a separate block, in a way assumed to be culturally inconsistent (e.g., men-submissive and women-dominant). Response latencies are compared across both the consistent and inconsistent testing blocks to give an indication as to which patterns of relational responding are more fluent and therefore more likely to be compatible with one's verbal history (e.g., Gavin, Roche, & Ruiz, 2008; O'Reilly, Roche, Ruiz, Tyndall, & Gavin, 2012).

The IRAP elaborates on the IAT procedure by assessing not only the existence of a pre-existing relation between stimuli, but also the specific way in which these stimuli are related. Specifically, the IRAP requires participants to respond quickly and accurately to a relation between two stimulus categories presented together in the context of specific relational terms (e.g., same/different) under alternating contingencies (e.g., men-submissive-*different* relative to men-submissive-*same*). Through a series of practice block pairs, participants are trained to a specific level of response fluency (typically requiring mean accuracies of over 80 % and individual response latencies of under 2000 ms). Once participants reach these criteria, their mean response latencies across critical test block pairs may be meaningfully compared. In this way, the IRAP is proposed to assess an individual's history of relating these stimuli, again resting on the core assumption that participants will respond more quickly and more accurately to relations that are consistent with their verbal history (Barnes-Holmes et al., 2011; Hughes, Barnes-Holmes, & Vahey, 2012).

The Function Acquisition Speed Test

An emerging behavior-analytic measure that adopts a slightly different approach to the assessment of verbal relations is known as the Function Acquisition Speed Test (FAST: O'Reilly et al., 2012). Essentially, the FAST shares a common overarching goal with the IRAP, in that it aims to assess whether natural verbal relations (i.e., those established through contact with the verbal community) impact laboratory-induced patterns of relational responding. However, where both the IAT and IRAP train participants to respond quickly and accurately with a series of practice/categorization blocks prior to testing, the FAST focuses on the rate of acquisition of those patterns themselves. That is, rather than investigating whether verbal histories facilitate or impede the speed at which classes of stimuli may be related together in a novel context, the FAST is concerned with the impact of history upon the differential rate at which these relations may be acquired.

In brief, the FAST procedure establishes functional equivalence classes between sets of stimuli by requiring participants to learn, via corrective feedback, a common response (e.g., a left or right key-press) when presented with individual stimuli on-screen. Participants are presented with two contrasting test blocks (one that establishes culturally-consistent response classes, and another that establishes culturally-inconsistent classes), with the primary metric being the difference in learning rate across the two blocks. Each block is comprised of four stimulus classes, with two classes sharing a left key-press response and two sharing a right key-press response (i.e., they are spatially and formally equivalent within the task). It is important to understand, however, that the parameters of the FAST can be modified to suit different populations or experimental requirements. It is crucial, for reasons to do with the behavior-analytic orientation of the test, that it not be conceived psychometrically, and that the various test parameters are always and only seen as functionally-understood features of a flexible general methodology.

In contrast to the IAT and IRAP, instructions on how to respond correctly are not given prior to the procedure, and no category/attribute labels or rules are provided at any point in the task. Rather, the establishment of a functional equivalence class is achieved exclusively through reinforcement (i.e., verbal feedback presented such as "Correct" and "Wrong"). As the FAST does not train response accuracy and latency in a practice phase (as in the IRAP), latency pressure is instead applied by restricting the response window to a relatively narrow time frame (usually 3000 ms), after which participants are presented with a timeout stimulus (e.g., "Wrong").

To date, the FAST paradigm has demonstrated utility in the assessment of both directly trained and derived equivalence relations established in the laboratory (O'Reilly, Roche,

Gavin, & Ruiz, 2013; O'Reilly et al., 2012). This current research aims to ascertain its effectiveness as a measure of natural stimulus relations, that is, those established through contact with a verbal community. Specifically, this research will test the sensitivity of the FAST to common gender stereotypes of women as stereotypically "feminine" (i.e., coordinated with traits such as empathic, submissive, gentle) and men as stereotypically "masculine" (i.e., coordinated with traits such as unemotional, assertive, dominant). Decades of social psychological research have evidenced the pervasiveness of the these gender stereotypes using traditional survey methods (e.g., Broverman, Vogel, Broverman, Clarkson, & Rosenkrantz, 1972; Eagly & Karau, 2002; Eagly & Steffen, 1984) as well as indirect measures such as the IAT (Rudman & Glick, 2001; Rudman & Kilianski, 2000). Though "complementary opposite" stereotypes such as these can at face value appear benevolent and/or innocuous (see Jost & Kay, 2005), their implications for gender inequality are well-documented, particularly in the workplace (e.g., Eagly & Karau, 2002; Heilman & Eagly, 2008; Rudman & Glick, 1999; Rudman, Moss-Racusin, Phelan, & Nauts, 2012; Rudman & Phelan, 2008). Simply put, as traditionally masculine traits tend to be those most socially valued in occupational settings (Eagly & Carli, 2007), the rigid coordination of these traits with men (and not women) is socially problematic. As such, their potential prevalence among a contemporary sample is worthy of investigation.

Modifications to the FAST Procedure

One important feature of the current research relates to a small number of procedural modifications to the FAST. The first involves the method of calculating the performance differential across the two key task blocks in the FAST. As stated previously, both the IAT and IRAP employ response latencies as their primary metric where, typically, latencies are measured from stimulus onset to first *correct* response, rather than to first response. A response correction procedure, in which incorrect responses are consequated by a prompt to produce the alternate (corrected) response is employed to artificially elongate reaction times (RTs) on error trials, in lieu of an older procedure in which no response correction was employed, but an arbitrary response time penalty was added to each RT recorded on error trials (i.e., the D_1 scoring algorithm as described by Greenwald, Nosek & Banaji, 2003). While this metric does not appear to undermine the predictive validity of the effects produced by either measure (see Lane, Banaji, Nosek, & Greenwald, 2007; Vahey et al., 2015), it is arguably somewhat opaque and metrics with more face validity can be easily employed (see below). The IAT and IRAP means of assessing RTs is reflective of a social-cognitive tradition dating back to Stroop (1935), in which RTs were used to index

processing effort rather than speed of responding. It is obvious, however, that the purpose of elongating raw RTs on error trials in the IAT (and IRAP) is precisely to reflect accuracy to some degree in the response time metric. In effect, a *fluency* measure to achieve this end would be both more transparent functionally, and more in line with behavioral tradition. Of greater concern here is the fact that the response correction technique works by definition as a form of punishment, rather than reinforcement. While it is employed by the IAT and IRAP ostensibly to help teach the participant to master the task block, it uses only negative feedback following response errors, and no feedback at all following correct responses. This approach to learning is counterintuitive, likely less effective than the use of positive reinforcement, and the delivery of punishers (a red X signaling an error) following responses has an unknown effect on error rates on subsequent trials (i.e., punishment should reduce accuracy, not increase it). In our research using the IAT and FAST over dozens of studies, we have noted anecdotally that error responses are roughly twice as likely to occur on a trial following a response correction procedure, compared to on those following positive feedback. The aim of the FAST, therefore, was to provide a simplistic, functionally transparent, implicit testing tool developed from the ground-up based on well-worn principles of behavior analysis (See Gavin et al., 2012; Ridgeway et al., 2010 for a more extended discussion of these issues).

In two published FAST studies to date, a more transparent index of learning speed has been employed; the difference in the natural log of the number of trials required to reach a specified response accuracy criterion across incompatible task blocks. However, this measure has proven to create inconsistent effects insofar as single errors in long response chains can radically alter the recorded trial requirements (e.g., trials required to produce 10 consecutive correct responses). Therefore, in the current study the slopes of the learning curves recorded across two finite (50-trial) task blocks were recorded as an alternative measure (i.e., a true correct response rate measure). Specifically, a cumulative record was created during each task block to provide both an absolute fluency rate for that block, but also to allow for a difference score in learning rates across blocks. The cumulative record allows us to assess and graph the rate of correct responding by plotting the number of correct responses as a function of unit time (Skinner, 1961). Though traditionally employed as a way to assess the rate of free operant responding, research on human participants has demonstrated its utility in assessing the rate of desired (e.g., correct) responses across a period of time (e.g., Hall, Hammond, Hirt, & Reiss, 2012; Rodgers & Iwata, 1991). This study, therefore, will explore the use of a cumulative record-type scoring procedure to assess and compare the rates of learning across the stereotype-consistent and stereotype-inconsistent FAST blocks.

A second important modification of previous FAST procedures relates to the simultaneous assessment of two pairs of stimulus relations, rather than just one. Previous FAST studies advocated the assessment of absolute or "non-relative" stimulus-stimulus relations (e.g., men-feminine relative to two neutral or previously unrelated stimulus classes, such as shapes-colors). However, such an approach may not be suitable in the current context, given the inherently relativistic nature of gender stereotypes. Both mainstream and RFT accounts posit that our very understanding of gender (i.e., roles, behaviors, and attributes) is rendered through the axiomatic construction of gender-as-binary, and of men and women as fundamental opposites (Connell, 2005; Hawkesworth, 1997; Kimmel, 2000; Roche & Barnes-Holmes, 2002). That is, it is not just that certain traits are male and others are female but, by virtue of the oppositional way in which gender is framed, male traits are also "not-female" and female traits are "not-male" (see also Prentice & Carranza, 2002). In this way, the concurrent juxtaposition of maleness with stereotypical masculinity and femaleness with stereotypical femininity in the FAST may be conceptually important in the analysis of gender stereotypes. Therefore, the current study elected to use a relative rather than non-relative approach here.

A third procedural modification relates to the removal of the "baseline" phases of the FAST. Previous research included two additional blocks, presented before and after the critical test blocks, which attempted to assess participants' baseline levels of response acquisition on two neutral or previously unrelated stimulus class training phases. However, given that the scoring procedure employed in the current study was focused on learning rate differentials across test blocks rather than raw rates of acquisition, baseline phases were not directly required here. Consequently, these blocks were removed from the protocol leaving just the two critical test blocks, as well as a short practice block to familiarize participants with the procedure.

The current study is thus the first investigation into the FAST's sensitivity to natural stimulus relations and, specifically, to common gender stereotypes. One "relative FAST" will be employed, which will contrast rates of learning across stereotype-consistent (i.e., men-masculine and women-feminine) and stereotype-inconsistent blocks (i.e., men-feminine and women-masculine). Given its previous utility in assessing implicit gender stereotypes (e.g., Rudman & Glick, 2001; Rudman & Kilianski, 2000), an IAT employing identical stimuli will also be employed as a positive control against which FAST scores can be compared. That is, FAST effects will be corroborated by the stereotype-consistent effects we expect to find on the IAT. In order to explore whether any meaningful relationships exist between FAST/IAT scores and more explicit forms of gender or sexual prejudice, a small number of self-report measures of sexism will also be employed. However, it should be noted that the precise conditions under which self-report measures are expected to correlate with implicit test effects is an ongoing issue for investigation within the field (see Nosek, 2007; Nosek, Greenwald, & Banaji, 2005). Thus, no specific hypotheses are made regarding the nature of any correlations between the FAST, IAT, and self-report questionnaires.

Method

Participants

Thirty undergraduate students (15 identified as female, 15 as male) from Maynooth University, aged between 18 and 26 participated in this study ($M_{age} = 20.47$, $SD = 1.634$). Participants were provided with an open-ended response format for gender information, so as not to impose a binarized forced-choice. Participation was voluntary and no remuneration was offered. Inclusion criteria included fluent English, normal or corrected-to-normal vision, and full use of both hands.

General Experimental Sequence

All experimental sessions were conducted one-to-one in individual experimental cubicles and all measures (including the self-reports) were administered on a laptop computer by a trained researcher. Before consenting to participate, all individuals were provided with a brief summary of the experiment's content and subject matter and were informed as to the general structure of the computer tasks and questionnaires. Once participants were comfortable with the nature of the experiment, written informed consent was obtained and they began the experiment. Participants were informed at the outset that they could cease participation at any time during the experiment. The general experimental sequence was as follows: FAST, self-report measures, and IAT. The FAST was employed prior to the self-report measures so as not to alert participants to the general purpose of the procedure (see also Nosek et al., 2005). The IAT, which was employed to corroborate FAST effects, was administered after the questionnaires in order to provide participants with a short break between computer-based tasks. Upon completion of all tasks, participants were fully debriefed regarding the study's aims and central thesis and were thanked for their time.

Materials

Self-Report Measures

Self-reported attitudes towards gender and sexuality were obtained from two questionnaires: the Modern Sexism Scale

(MS: Swim, Aikin, Hall, & Hunter, 1995) and the Heteronormativity Attitudes and Beliefs Scale (HABS; Habarth, 2015).

Modern Sexism Scale The MS is a 10-item scale assessing beliefs about women and gender, comprising two 5-item subscales. The first is intended to assess more traditional anti-women sentiments (e.g., "It is more important to encourage boys than to encourage girls to participate in athletics."), and the second examines more subtle or contemporary sexist attitudes (e.g., "Over the past few years, the government and news media have been showing more concern about the treatment of women than is warranted by women's actual experiences."). Items are scored on a Likert-scale from 1 (strongly disagree) to 5 (strongly agree), with possible scores ranging from 10-50.

Heteronormativity Attitudes and Beliefs Scale The HABS is a 16-item questionnaire assessing the beliefs that gender is binary (i.e., that people can or must be either male or female) and that heterosexuality is a natural feature of male-female relations. Two 8-item subscales assess, firstly, "gender-as-binary" beliefs (e.g., "All people are either male or female") and, secondly, attitudes around normative sexual behavior (e.g., "There are particular ways that men should act and particular ways that women should act in relationships"). Items are scored on a 7-point Likert scale from "strongly disagree" to "strongly agree", allowing a scoring range of 16-112.

FAST

The version of the FAST utilized in the current experiment was developed with the computer programming software Livecode™ and was administered to participants on a laptop computer with a 13" screen (1024 × 768 pixel resolution). As previously mentioned, the FAST protocol in the current experiment deviated somewhat from previous published FAST studies (e.g., O'Reilly et al., 2012; 2013) by consisting of just one practice block and two test blocks (baseline blocks were omitted). All blocks contained four distinct stimulus categories, each consisting of 2–4 exemplars (see Table 1). For the practice blocks, stimulus categories consisted of common, everyday words that were deemed neutral in terms of their evaluative functions (i.e., the categories "fruits", "animals", "body parts", and "clothes"). The test blocks were identical in terms of the format and stimuli included, differing only in terms of the reinforcement contingencies applied. Specifically, test block 1 reinforced one shared key-press response for "men" and "masculine traits" and another for "women" and "feminine traits". That is, corrective feedback reinforced responding in a stereotype-consistent manner. These contingencies were reversed for test block 2 (blocks 1 and 2 were

Table 1 Stimuli used in the gender stereotype FAST and IAT

Target 1 Men	Target 2 Women	Attribute 1 Stereotypically masculine traits	Attribute 2 Stereotypically meminine traits
Male	Female	Dominant	Nurturing
		Unemotional	Gentle
Man	Woman	Competitive	Bitchy
		Aggressive	Emotional

Note: The word "bitchy" in the Irish vernacular is generally assumed to denote someone who readily makes spiteful or malicious remarks. It is likely this usage is culturally specific, and should be noted by readers for whom English is not their first language

presented in a randomized order across participants). The reader is reminded, at this point, that no instructions were provided to the participants regarding how they should respond; response classes were shaped entirely using a FR1 reinforcement schedule (i.e., trial-by-trial feedback).

Before commencing, participants were presented with the following instructions outlining the general nature of the task:

> In the following section, your task is to learn which button to press when a word appears on screen. IMPORTANT: During this phase you should press only the Z key or the M key. Please locate them on the keyboard now. To help you learn you will be provided with feedback telling you if you are right or wrong. If you have any questions please ask the researcher now. Press any key when you are ready to begin.

After participants had pressed a random key, they were then presented with a practice block consisting of 16 trials, followed by two test blocks each consisting of 50 trials. Identical instructions to the ones mentioned above were provided before each block. Each stimulus was presented for 3 seconds, with all stimuli presented in black, 48-type font in the center of the screen. The presentation of each stimulus was followed by a 250-ms intertrial interval. The shared response function (i.e., a common key-press response) in each block was reinforced via corrective feedback, which was presented after the participant's key-press response. The feedback ("Correct" or "Wrong") was presented on-screen for a period of 1.5 seconds in the center of the screen in red, 48-type font. If no response was emitted from participants within the 3-second response window (i.e., no key was pressed), a time-out response ("Wrong") was presented in the center of the screen, again in red, 48-point font. The order of presentation of the two test blocks, and of the stimuli within them, was randomized. Upon completion of the task, participants were provided with a brief message letting them know the task was over and thanking them for taking part.

IAT

The IAT in the current experiment adhered to the standard format (see Greenwald et al., 1998) and was programmed in Inquisit^TM software. Identical stimuli to those employed in FAST were used (Table 1). In brief, the task comprised five separate phases, presented in the following order. Firstly, participants completed a target discrimination task, in which they were required to categorize exemplars from the two target categories according to their appropriate labels (e.g., sorting the word "man" into the category *masculine*). Similarly, the second phase of the IAT required participants to complete a near identical task, this time for the attribute categories (e.g., sorting the word "nurturing" into the category *feminine*). In both tasks, each trial presented an individual exemplar (e.g., "dominant") in the center of the screen and the two category labels (*masculine* or *feminine*) at the top left and right of the screen. In phase 3, participants were required to sort both targets and attributes simultaneously, with both target and category labels presented together at the top left and right of the screen. The first task required participants to sort stimuli in an manner presumed to be "consistent" with their verbal histories, that is, a common key press for words that are either *male* or *masculine* and another for those either *female* or *feminine*. Phase 4 then repeated the target discrimination task, this time requiring participants to sort exemplars in an "inconsistent" way (e.g., "man" into the category *female*). Lastly, phase 5 was identical to phase 3, but the target/label pairings were reversed (e.g., words that are *female* or *masculine* share a response).

In all phases of the IAT, participants were required to categorize exemplars by pressing keys on either the left or right of the keyboard, that is, to press "e" for words belonging to the category to the left of the screen and "i" for words to the right of the screen. No feedback was provided for correct responses, while incorrect responses resulted in a red 'X' in the center of the screen. All stimuli and error messages remained on-screen until a correct response was provided, with a 250-ms interval between trials.

Data Processing

The primary metric of the FAST was learning rate, derived from the participant's cumulative record of correct responses. More specifically, a cumulative record was generated for both the "consistent" and "inconsistent" blocks by plotting the number of correct responses (the response being reinforced) per block as a function of time. The rate of learning corresponds to the slope of this function (calculated using the SLOPE function in excel), with a higher slope indicative of a faster rate. Accuracy and response latency (in seconds) were recorded by the software on a trial-by-trial basis. Time for the

entire block was cumulated and included the 250-ms intertrial intervals. The learning rate differential (i.e., the difference in the slope coefficients of consistent and inconsistent blocks) was quantified by subtracting the slope of the consistent block from that of the inconsistent block. A positive raw difference score therefore reflected a men-masculine/women-feminine effect, while a negative score reflected a women-masculine/men-feminine effect. Given the widely-noted issues inherent in using raw difference (see Edwards & Parry, 1993), learning rate differentials were standardized using standard errors.

The IAT score was calculated using the popular D_1 scoring algorithm, a metric quantifying response latencies based on Cohen's *d* (see Greenwald et al., 2003). Scoring practices and interpretation of IAT scores in the current experiment were identical to that outlined by Greenwald and colleagues, and thus will not be discussed in detail here. However, for ease of interpretation and comparison, it should be noted that a positive *D* score similarly reflects a men-masculine/women-feminine effect, a negative score indicates a women-feminine/men-masculine effect, and that scores ranged from -2 to +2, as standard.

Results

Function Acquisition Speed Test

Learning Rate

Learning rates (i.e., slope values) for both blocks of the FAST are presented in Table 2. As can be seen in Table 2, 29 of 30 participants demonstrated faster rates of learning on the stereotype-consistent relative to the stereotype-inconsistent training block, as expected (i.e., their raw difference score was positive rather than negative). A Wilcoxon signed-rank test revealed the overall FAST effect (i.e., the difference between the two blocks) to be significant, $z = -4.39$, $p < 0.001$, with a large effect size ($r = .8$). With regard to any potential order effects, a Mann-Whitney *U* test revealed no significant differences in latency differentials based on the order in which FAST blocks were administered, $U = 106$, $z = -.250$, $p < .082$.

Sample Cumulative Record: Visual Analysis

Figure 1 illustrates a sample cumulative record for participant 1. With the exception of one participant, faster (i.e., steeper) learning rates were observed for the stereotype-consistent block relative to the stereotype-inconsistent block. As evidenced by the graph, response patterns in early trials (the first 25 seconds for this participant) are nearly identical across the two blocks. After this point, learning rates are noticeably steeper for the stereotype-consistent relative to the

Table 2 FAST and IAT data

Participant	Gender	Stereotype-consistent block Slope of learning curve	Stereotype-inconsistent block Slope of learning curve	Raw difference	Standardized difference	IAT score
1	F	0.83	0.67	.16	-0.12	1.24
2	M	0.75	0.67	.08	-0.75	1.14
3	F	0.6	0.44	.16	-0.12	-.49
4	M	0.57	0.29	.28	0.83	1.08
5	F	0.79	0.57	.22	0.35	.92
6	M	0.63	0.55	.08	-0.75	.71
7	F	0.79	0.4	.39	1.69	1.29
8	M	0.75	0.56	.19	0.12	.48
9	M	0.75	0.4	.35	1.38	.19
10	F	0.77	0.62	.15	-0.2	1.03
11	F	0.76	0.32	.44	2.09	1.18
12	F	0.82	0.63	.19	0.12	.79
13	F	0.78	0.62	.16	-0.12	.98
14	F	0.77	0.56	.21	0.27	.75
15	M	0.67	0.58	.09	-0.67	
16	F	0.63	0.26	.37	1.54	.74
17	M	0.72	0.46	.26	0.67	.57
18	M	0.72	0.47	.25	0.59	.59
19	M	0.68	0.44	.24	0.51	1.14
20	M	0.61	0.44	.17	-0.04	.98
21	F	0.45	0.66	-.21	-3.04	.78
22	F	0.88	0.62	.26	0.67	1.10
23	F	0.77	0.72	.05	-0.99	.93
24	F	0.74	0.68	.06	-0.91	.38
25	M	0.77	0.64	.13	-0.36	.56
26	M	0.83	0.79	.04	-1.07	.52
27	F	0.66	0.51	.15	-0.2	.63
28	M	0.64	0.49	.15	-0.2	.87
29	M	0.77	0.63	.14	-0.28	
30	M	0.72	0.67	.05	-0.99	.90
Means		0.72	0.55	0.17	0.001	0.79

Note: Participants 15 and 29 did not complete an IAT due to time constraints

stereotype-inconsistent block. Though figure 1 is presented here only as an example, learning trajectories for the majority of participants followed a similar pattern. Presumably, participants made a roughly equal number of mistakes during early trials given that the reinforcement contingencies had yet to have a measureable impact on behavior. As the blocks progressed, however, an increasing number of responses were correct, but more so during the stereotype-consistent block.
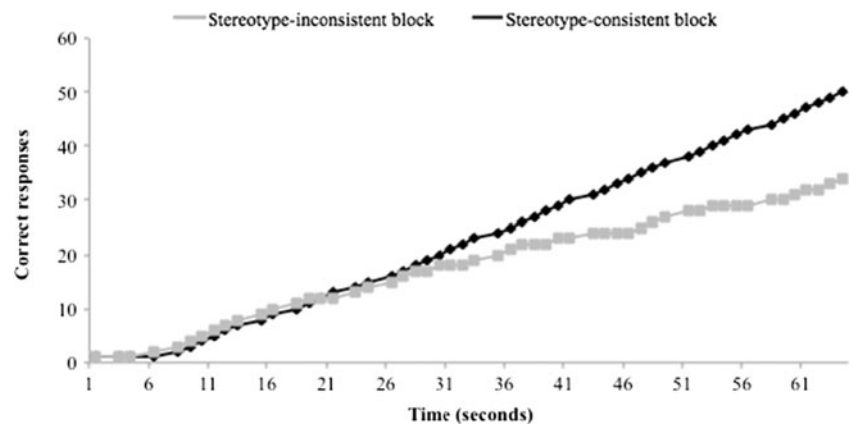
Self-Report Measures

Self-reported sexism was low in the current sample, with scores ranging from 10-29 ($M = 16.10$, $SD = 4.74$).

Heteronormativity scores were more varied, ranging from 16-105 ($M = 35.8$, $SD = 23.93$), though scores tended to cluster around the low end.

Implicit Association Test

As seen in Table 2, 27 of the 28 participants who completed an IAT demonstrated a positive IAT effect (i.e., response latencies were shorter for the stereotype-consistent relative to the stereotype-inconsistent block). Single-sample t-tests revealed the overall IAT effect to be significant against zero, $t(27) = 11.17$, $p < .001$.

Fig. 1 Sample cumulative record for participant 1. Note: steeper graphs represent the rate at which participants learned to respond correctly on the discrimination tasks in each training block. A one-point vertical increment on the graph corresponds to one correct response. A horizontal increment reflects the passage of a 3-s trial without a correct response (i.e., a missed response or an incorrect response)

Measure Comparisons

No significant correlations were found between self-reported sexism and heteronormativity and scores on either the FAST (standardized difference score) or the IAT ($ps > .5$) using Pearson's product moment correlation. Additionally, while the IAT and the FAST produced very consistent co-varying effects (i.e., the effect was in the same direction for all participants with the exception of participant 3 and participant 21; see Table 2), no correlations were found between the IAT and FAST scores ($p > .5$). More consideration of the relationship between these two measures can be found in the discussion.

Gender Differences

Independent sample t-tests revealed no difference between male and female participants on either the FAST, $t(28) = -.369$, $p > .5$, or the IAT, $t(28) = -.481$, $p > .5$. Gender differences were found for both of the explicit measures, however, with men scoring significantly higher than women on the MS, $t(28) = 2.262$, $p = .001$, and the HABS, $t(28) = 2.152$, $p = .001$. Effect sizes for both tests were large ($d > .8$).

Discussion

The current study expands on previous research (O'Reilly et al., 2012; 2013) by demonstrating the utility of the FAST procedure in assessing the nature of verbal stimulus-stimulus relations. Specifically, this procedure proved to be sensitive to pervasive and common gender stereotypes of women as stereotypically "feminine" (e.g., emotional, nurturing) and men as stereotypically "masculine" (e.g., dominant, aggressive). Of the 30 individuals who participated in this study, all demonstrated effects in the expected direction insofar as learning rates were significantly faster for the stereotype-consistent discrimination block relative to the stereotype-inconsistent block. Similar effects were observed on an IAT containing

identical stimuli, which was employed here as a positive control. Specifically, 27 of the 28 participants who completed an IAT demonstrated faster response latencies for the consistent relative to the inconsistent block, thereby corroborating effects found on the FAST.

The current findings support the use of a stimulus equivalence or broader derived stimulus relations paradigm in the assessment of natural verbal relations. In so doing, this study further supports the idea posed in the Watt et al. (1991) procedure that differential rates of equivalence class formation may be used to assess distal learning histories. However, in slight contrast to the Watt et al. procedure and measures such as the IAT or IRAP, these findings suggest that this can be achieved using a very quick (two test blocks taking approx. five minutes to complete) and easy-to-administer procedure. In addition, the use of a cumulative record as a functionally understood metric of learning rate appears to be a sensitive scoring technique and one consistent with traditional means of assessing learning rates in the experimental analysis of behavior.

Given that each measure produced significant and robust effects in the expected direction (with the exception of two participants), it is reasonable to infer that both the IAT and the FAST reliably measured verbal histories around gender. The lack of direct correlation between the two measures, however, is worthy of consideration. While spurious or absent correlations between two implicit measures is not uncommon (Bosson, Swann, & Pennebaker, 2000; Hussey, Daly, & Barnes-Holmes, 2015), the paradigmatic differences across the two measures may point to the effects of even slight alterations to procedural or quantification methods. More specifically, it could be possible that the IAT and FAST index different features of response strength, or are differentially sensitive to different products of a relational history (i.e., relative speed of responding versus relative rate of learning across fixed discrimination blocks). In this way, while latency and learning rate differentials should roughly co-vary (as observed), they may not necessarily be linearly related. While it remains to be seen if these different metrics are of more or less use in terms

of their predictive and/or construct validity, the lack of correlation produced by two putatively similar procedures nonetheless opens up further lines of inquiry regarding the most precise means of indexing stimulus-stimulus relations in the vernacular (see also Hughes et al., 2012 for a review of some alternative or non-latency based scoring metrics).

The absence of a linear relationship between the implicit measures and the questionnaire scores is also not entirely unexpected. As mentioned in the introduction, the conditions under which explicit and implicit measures correlate is an ongoing area of investigation within the field of implicit testing (see Nosek, 2007). The Relational Elaboration and Coherence (REC) model, for instance, accounts for explicit-implicit variation by drawing a pragmatic conceptual distinction between "brief and immediate" relational responses (BIRRs), or those emitted most readily in a time-pressured environment, and "extended and elaborated" relational responses (EERRs), or those emitted in more open-ended time periods following deliberation (see Barnes-Holmes et al., 2011). According to this model, BIRRs and EERRs may align under certain conditions (e.g., matching shapes to their corresponding words) but may potentially diverge in certain contexts that require greater deliberation or abstraction (e.g., political views), or where certain patterns of responding are more socially desirable than others (e.g., liberal versus conservative views around gender; but also see O'Reilly, Roche, & Cartwright, 2014). Although an exploration into the relationship between self-reported sexism and implicit gender stereotypes was not the goal of this study, it is plausible that the lack of correlation reflects a tendency among participants to respond differently, and potentially more liberally, to gender-relevant stimuli in the absence of time pressure. To assess the social relevance of this divergence, future research could perhaps investigate the relative abilities of self-reported beliefs and implicit stereotypes to predict some discriminatory behavior (e.g., hiring prejudices), as has been done elsewhere (e.g., Rudman & Glick, 2001).

In terms of the gender stereotypes identified in the study, their strength and prevalence in the current sample cohere with the findings of previous IAT research in this area (Rudman & Glick, 2001; Rudman & Kilianski, 2000). That is, participants demonstrated a strong and significant bias toward coordinating men with traditionally masculine (but not feminine) traits and women with traditionally feminine (but not masculine) traits. The separable divide between personality traits on the FAST and IAT (i.e., the strong effects observed) therefore suggests an apparent persistence of gender stereotypes in modern society and, more broadly, elucidates the binary way in which gender is constructed. That is, it evinces that that which is male is also "not-female", and vice versa. Such an interpretation of these findings is consistent with mainstream social psychological and feminist theories concerning the socialization of gender-as-binary (e.g., Connell, 2005; Eagly & Steffen, 1984; Hawkesworth, 1997; Kimmel, 2000).

Future research should now explore whether the effects generated by the FAST constitute a meaningful variable in the analysis of social behavior. As previously stated, gender stereotypes of women as gentle/nurturing and men as dominant/competitive have noted implications for gender equality (see Eagly & Karau, 2002; Rudman et al., 2012). To date, research has demonstrated the utility of gender-stereotype IAT procedures and the latency-based d score in predicting more overt forms of gender-based discrimination. For instance, Rudman and Glick (2001) demonstrated that effects on an IAT using similar stimuli predicted negative social evaluations of female job applicants in occupational settings. Similar techniques could be employed as a way to validate the FAST procedure and scoring technique and to assess its predictive utility for prejudicial behavior. For example, it could be useful to compare FAST scores between two groups who should theoretically differ in the extent to which they endorse traditional gender stereotypes (e.g., between self-identified feminists and individuals with known anti-women biases). Similarly, it would be interesting to investigate whether performance on the FAST is predictive of behavior on some socially meaningful task, such as the Rudman and Glick hiring preference paradigm.

Concerning the procedural modifications to the FAST, findings of this study support the use of the "relativistic" paradigm for assessing social biases, although its necessity still warrants investigation. As stated in the introduction, previous research (O'Reilly et al., 2013) noted the FAST's potential utility as a measure of "absolute" or non-relative biases (e.g., of "men" and "good", relative to two unrelated stimulus categories). The current study, however, opted to assess gender stereotypes of men and women relativistically (i.e., in relation to each other) in a single FAST. The rationale for this was, in part, based on previous research into common gender stereotypes, which the current study was partly replicating, as well as on IAT findings suggesting an important role for counter-categories in specifying the dimension of comparison (e.g., women/men relative to women/objects; Rudman & Mescher, 2012). That is, if one is interested in assessing gender stereotypes, the juxtaposition of gender categories with their corresponding traits may be an important way to bring a particular history to bear on the discrimination task. Put simply, it is possible that learning to coordinate women with certain traits in this study was facilitated by the concurrent coordination of men with others, as this may have specified the two dimensions of comparison (i.e., men, women, male traits, and female traits). However, such an interpretation warrants further investigation into the relative sensitivity and predictive validity of relative versus non-relative FAST procedures.

One potential limitation of this study concerns the order in which the FAST and IAT were presented. The FAST was

always presented first and the IAT last, given that the IAT was employed solely as a positive control against which the FAST effects could be compared. Although the intention of this study was not to engage in a direct IAT/FAST comparison or to assess the superior construct/predictive validity of either measure, it is acknowledged that counter-balancing the order of the two would have allowed for a more detailed comparison of measure effects. The undergraduate sample employed here may also constitute a limitation, given that a study examining contemporary gender stereotypes should employ a more representative sample (e.g., in terms of age, educational background, socioeconomic status, etc.). However, it should be remembered that the purpose of the current study was not so much to survey the existence of these verbal categories but to test the utility of the newly revised FAST procedure. In effect, the representativeness of the sample employed here does not detract from an in-principle demonstration of the FAST as a viable procedure for use with real-world stimulus categories. Nevertheless, the relative impact of different demographic variables on test outcomes should be explored directly in future research aiming to survey stereotypes in the general population.

Summary

The current study provides empirical support for a functionally understood implicit test style method for assessing verbal relations, based directly on the Watt et al. (1991) stimulus equivalence paradigm. Though it should be noted that the FAST is unlikely to provide as extensive or detailed an assessment of relational responding histories as the IRAP, the FAST may nevertheless have certain pragmatic advantages. First, it does not require any participant pre-training, and does not appear to be a difficult format for participants to master (e.g., it had a zero participant attrition rate in the current study). Second, that the FAST requires little explanation, instruction, or practice for participants means it is both quick (under 5 minutes) and easy-to-administer. Third, the scoring technique that was employed here was novel in the field of implicit testing but represents a conceptually clear metric of learning rate differentials that is based on time-tested behavior-analytic principles. As such, the FAST may represent a measure of verbal relations that may prove useful to behavioral researchers in assessing the emergence of trained discriminations, derived relations, and real-world verbal relations more generally.

References

Barnes-Holmes, D., Barnes-Holmes, Y., Stewart, I., & Boles, S. (2011). A sketch of the implicit relational assessment procedure (IRAP) and the relational elaboration and coherence (REC) model. *The Psychological Record, 60*(3), 527–542.

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*(4), 631–643. doi:10.1037//0022-3514.79.4.631.

Broverman, I. K., Vogel, S. R., Broverman, D. M., Clarkson, F. E., & Rosenkrantz, P. S. (1972). Sex-role stereotypes: A current appraisal. *Journal of Social Issues, 28*(2), 59–78. doi:10.1111/j.1540-4560.1972.tb00018.x.

Connell, R. W. (2005). Hegemonic masculinity: Rethinking the concept. *Gender & Society, 19*(6), 829–859. doi:10.1177/0891243205278639.

Eagly, A. H., & Carli, L. L. (2007). *Through the labyrinth: The truth about how women become leaders*. Boston, MA: Harvard Business School Press.

Eagly, A. H., & Karau, S. J. (2002). Role congruity theory of prejudice toward female leaders. *Psychological Review, 109*(3), 573–598. doi:10.1037//0033-295X.109.3.573.

Eagly, A. H., & Steffen, V. J. (1984). Gender stereotypes stem from the distribution of women and men into social roles. *Journal of Personality and Social Psychology, 46*(4), 735–754. doi:10.1037/0022-3514.46.4.735.

Edwards, J. R., & Parry, M. E. (1993). On the use of polynomial regression equations as an alternative to difference scores in organizational research. *The Academy of Management Journal, 36*(6), 1577–1613. doi:10.2307/256822.

Gavin, A., Roche, B., Ruiz, M. R., Hogan, M., & O'Reilly, A. (2012). A behavior analytically modified implicit association test for measuring sexual categorization of children. *The Psychological Record, 62*(1), 55–68.

Gavin, A., Roche, B., & Ruiz, M. R. (2008). Competing contingencies over derived relational responding: A behavioral model of the Implicit Association Test. *The Psychological Record, 57*, 427–441.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology, 74*(6), 1464–1480.

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.

Habarth, J. M. (2015). Development of the heteronormative attitudes and beliefs scale. *Psychology & Sexuality, 6*(2), 166–188. doi:10.1080/19419899.2013.876444.

Hall, S. S., Hammond, J. L., Hirt, M., & Reiss, A. L. (2012). A "learning platform" approach to outcome measurement in fragile X syndrome: a preliminary psychometric study. *Journal of Intellectual Disability Research, 56*(10), 947–960. doi:10.1111/j.1365-2788.2012.01560.x.

Hawkesworth, M. (1997). Confounding gender. *Signs, 22*(3), 649–685.

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (2001). *Relational frame theory: A post-Skinnerian account of human language and cognition*. New York, NY: Kluwer Academic/Plenum Publishers.

Heilman, M. E., & Eagly, A. H. (2008). Gender stereotypes are alive, well, and busy producing workplace discrimination. *Industrial and Organizational Psychology, 1*(4), 393–398.

Hughes, S., & Barnes-Holmes, D. (2016). Relational Frame Theory: The basic account. In S. C. Hayes, D. Barnes-Holmes, R. D. Zettle, & A. Biglan (Eds.), *Handbook of Contextual Behavioral Science*. New York: Wiley-Blackwell.

Hughes, S., Barnes-Holmes, D., & Vahey, N. (2012). Holding on to our functional roots when exploring new intellectual islands: A voyage through implicit cognition research. *Journal of Contextual Behavioral Science, 1*(1–2), 17–38. doi:10.1016/j.jcbs.2012.09.003.

Hussey, I., Barnes-Holmes, D., & Barnes-Holmes, Y. (2015a). From relational frame theory to implicit attitudes and back again: Clarifying the link between RFT and IRAP research. *Current Opinion in Psychology, 2*, 11–15. doi:10.1016/j.copsyc.2014.12.009.

Hussey, I., Daly, T., & Barnes-Holmes, D. (2015). Life is good, but death ain't bad either: Counter-intuitive implicit biases to death in a normative population. *The Psychological Record*, Advance online publication. http://doi.org/10.1007/s40732-015-0142-3.

Jost, J. T., & Kay, A. C. (2005). Exposure to benevolent sexism and complementary gender stereotypes: Consequences for specific and diffuse forms of system justification. *Journal of Personality and Social Psychology, 88*(3), 498–509. doi:10.1037/0022-3514.88.3.498.

Kimmel, M. S. (2000). *The gendered society*. New York, NY: Oxford University Press.

Lane, K. A., Banaji, M. R., Nosek, B. A., & Greenwald, A. G. (2007). Understanding and using the implicit association test: IV. What we know (so far) about the method. In B. Wittenbrink & N. Schwartz (Eds.), *Implicit measures of attitudes* (pp. 59–102). New York, NY: Guildford Press.

Nosek, B. A. (2007). Implicit–explicit relations. *Current Directions in Psychological Science, 16*(2), 65–69.

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2005). Understanding and using the implicit association test: II. Method variables and construct validity. *Personality & Social Psychology Bulletin, 31*(2), 166–180. doi:10.1177/0146167204271418.

O'Reilly, A., Roche, B., Gavin, A., & Ruiz, M. R. (2013). A function acquisition speed test for equivalence relations (FASTER). *The Psychological Record, 63*, 707–724.

O'Reilly, A., Roche, B., Ruiz, M., Tyndall, I., & Gavin, A. (2012). The function acquisition speed test (FAST): a behavior analytic implicit test for assessing stimulus relations. *The Psychological Record, 62*(3), 507–528.

O'Reilly, A., Roche, B., & Cartwright, A. (2014). Function over form: A modern behaviour-analytic approach to implicit attitudes. In J. Zin

(Ed.), *Exploring implicit cognition: Learning, memory, and social-cognitive processes*. IGI Global: Hershey, PA.

Prentice, D. A., & Carranza, E. (2002). What women and men should be, shouldn't be, are allowed to be, and don't have to be: The contents of prescriptive gender stereotypes. *Psychology of Women Quarterly, 26*(4), 269–281. doi:10.1111/1471-6402.t01-1-00066.

Ridgeway, I., Roche, B., Gavin, A., & Ruiz, M. R. (2010). Establishing and eliminating implicit association test effects in the laboratory: Extending the behavior-analytic model of the implicit association test. *European Journal of Behavior Analysis, 11*, 133–150.

Roche, B., & Barnes-Holmes, D. (2002). Human sexual arousal: A modern behavioral approach. *The Behavior Analyst Today, 3*(2), 145–154.

Roche, B., & Dymond, S. (2013). *Advances in relational frame theory: Research and application*. Oakland: New Harbinger Publications.

Rodgers, T. A., & Iwata, B. A. (1991). An analysis of error-correction procedures during discrimination training. *Journal of Applied Behavior Analysis, 24*(4), 775–781. doi:10.1901/jaba.1991.24-775.

Rudman, L. A., & Glick, P. (1999). Feminized management and backlash toward agentic women: The hidden costs to women of a kinder, gentler image of middle managers. *Journal of Personality and Social Psychology, 77*(5), 1004–1010. doi:10.1037/0022-3514.77.5.1004.

Rudman, L. A., & Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues, 57*(4), 743–762.

Rudman, L. A., & Kilianski, S. E. (2000). Implicit and explicit attitudes toward female authority. *Personality and Social Psychology Bulletin, 26*(11), 1315–1328. doi:10.1177/0146167200263001.

Rudman, L. A., & Mescher, K. (2012). Of animals and objects: Men's implicit dehumanization of women and likelihood of sexual aggression. *Personality and Social Psychology Bulletin, 38*(6), 734–746. doi:10.1177/0146167212436401.

Rudman, L. A., Moss-Racusin, C. A., Phelan, J. E., & Nauts, S. (2012). Status incongruity and backlash effects: Defending the gender hierarchy motivates prejudice against female leaders. *Journal of Experimental Social Psychology, 48*(1), 165–179. doi:10.1016/j.jesp.2011.10.008.

Rudman, L. A., & Phelan, J. E. (2008). Backlash effects for disconfirming gender stereotypes in organizations. *Research in Organizational Behavior, 28*, 61–79. doi:10.1016/j.riob.2008.04.003.

Skinner, B. F. (1961). *Cumulative record*. New York: Appleton-Century-Crofts.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643–662. doi:10.1037/h0054651.

Swim, J. K., Aikin, K. J., Hall, W. S., & Hunter, B. A. (1995). Sexism and racism: Old-fashioned and modern prejudices. *Journal of Personality and Social Psychology, 68*(2), 199–214. doi:10.1037/0022-3514.68.2.199.

Vahey, N. A., Nicholson, E., & Barnes-Holmes, D. (2015). A meta-analysis of criterion effects for the implicit relational assessment procedure (IRAP) in the clinical domain. *Journal of Behavior Therapy and Experimental Psychiatry, 48*, 59–65. doi:10.1016/j.jbtep.2015.01.004.

Watt, A., Keenan, M., Barnes, D., & Cairns, E. (1991). Social categorization and stimulus equivalence. *The Psychological Record, 41*(1), 33–50.