International Journal of Psychology and Psychological Therapy, 2019, 19, 3, 323-336

Exploring the Use of Pictures of Self and Other in the IRAP: Reflecting upon the Emergence of Differential Trial Type Effects

Deirdre Kavanagh
Ghent University, Belgium
Nele Matthyssen
Thomas More Hogeschool, Antwerp, Belgium
Yvonne Barnes-Holmes, Dermot Barnes-Holmes, Ciara McEnteggart
Ghent University, Belgium
Roberta Vastano
University of Miami, USA

Abstract

The Implicit Relational Assessment Procedure (IRAP) assesses the relative strength of derived relational responding. A growing body of IRAP research has focused on assessing verbal relations pertaining to the self and others. This preliminary study sought to determine the feasibility of using matched pictures of self and of others across two IRAPs (*N*= 32). Both the self- and other-IRAPs also presented pictures of pens as the contrast category. The results of the IRAPs were broadly consistent with common-sense expectations. That is, participants confirmed more readily than they denied that a picture of a face was a face and that a picture of a pen was a pen. They also denied more readily than confirmed that a picture of a pen was a face and that a picture of a face was a pen. No significant differences in the sizes of the individual trial type effects, or differences among those effects, emerged between the two (self and other) IRAPs. However, two key differential trial type effects did emerge for both IRAPs, which relate directly to recent and on-going conceptual developments surrounding the IRAP and the analysis of the dynamics of arbitrarily applicable relational responding in general. These developments are considered and discussed in detail toward the end of the article.
Key words: Relational Frame Theory, IRAP, differential trial type effects, self.

How to cite this paper: Kavanagh D, Matthyssen N, Barnes-Holmes Y, Barnes-Holmes D, McEnteggart C, & Vastano R (2019). Exploring the Use of Pictures of Self and Other in the IRAP: Reflecting upon the Emergence of Differential Trial Type Effects. *International Journal of Psychology & Psychological Therapy, 19, 3,* 323-336.

Novelty and Significance

What is already known about the topic?

- The Implicit Relational Assessment Procedure (IRAP) has been used to explore responding to self versus other across a range of domains, such as self-esteem and perspective-taking.
- Typically, the IRAPs employed in these studies contain words or terms that refer to self (e.g., "I") and others (e.g., "others").
- Although progress has been made using these types of stimuli, the potential for ambiguity for participants in interpreting the correct person reference in the IRAP remains. For example, when "I" appears on-screen it may be interpreted as referring to the computer rather than the participant.

What this paper adds?

- The current study is novel in that it sought to determine if using pictures of the self versus another is feasible in the IRAP.
- Specifically, using a picture of each participant and a picture of another may serve to reduce the ambiguity noted above regarding the interpretation of who is being referred to in IRAP trials (especially in reference to "I" the participant).
- The current research is also relatively novel in that some of the effects obtained are interpreted, albeit in a post-hoc fashion, in the light of a recent model of differential IRAP effects that have been observed across a range of IRAP studies.

Correspondence: Deirdre Kavanagh, Department of Experimental, Clinical, and Health Psychology, Ghent University, 9000 Ghent, Belgium. Email: deirdre.kavanagh@ugent.be. Acknowledgements: The data for the current manuscript was collected and prepared with the support of the FWO Type I Odysseus Program at Ghent University, Belgium; This article was prepared with the support of an Odysseus Group 1 Grant awarded to the fourth author by Flanders Science Foundation (FWO).

Behavior-analytic researchers have developed increasingly complex accounts of human language and cognition in terms of derived stimulus relations. A particularly rich vein of research in this regard is known as relational frame theory (RFT; Hayes, Barnes-Holmes, & Roche, 2001; for a recent review, see Hughes & Barnes-Holmes, 2016). Much of the work conducted in this area has focused on demonstrating specific patterns of derived relational responding in a dichotomous manner. That is, participants were typically trained and tested to determine if they produced a predicted pattern, but the research rarely focused on the relative strength of a pattern once it was observed. There have been recent calls, however, for analyses that focus on the relative strength or persistence of derived relational responding, rather than simply its presence versus absence (e.g., Barnes-Holmes, Barnes-Holmes, Hussey, & Luciano, 2016). In an attempt to develop methodologies for assessing the relative strength of derived relational responding, researchers have explored alternative methodologies, such as the Implicit Relational Assessment Procedure (IRAP; e.g., Barnes-Holmes, Hayden, Barnes-Holmes, Stewart, & Boles, 2008; Hussey, Barnes-Holmes, & Barnes-Holmes, 2015).

The IRAP requires participants to emit two opposing patterns of relational responding, and the ease with which one pattern may be emitted relative to the other provides a measure of response strength. Specifically, the procedure typically presents label and target stimuli (e.g., the label word "puppy" with the target word "pleasant") and requires participants to confirm or disconfirm the relational coherence between these stimuli (i.e., select "True" on coherent trials such as *Puppy-Pleasant* and "False" on incoherent trials such as *Spider-Pleasant*). An IRAP typically comprises four trial types (e.g., *Puppy-Positive*, *Puppy-Negative*, *Spider-Positive*, and *Spider-Negative*) that are generally analyzed independently in terms of the difference in response latencies between responding that is deemed consistent (coherent) versus inconsistent (incoherent) with a participant's verbal history. In general, response latencies are expected to be shorter during blocks of trials that require history-consistent versus history-inconsistent responding.

Some IRAP research has focused on assessing verbal relations pertaining to self versus others. This work has included areas such as self-esteem and perspective-taking. For example, Remue, De Houwer, Barnes-Holmes, Vanderhasselt, and De Raedt (2013) assessed what they referred to as actual and ideal self-esteem across two IRAPs. In the actual self-esteem IRAP, the label stimuli comprised statements such as "I am" or "I am not", while the ideal self-esteem IRAP presented "I want to be" and "I don't want to be", along with either positive or negative target stimuli (e.g., "happy" or "guilty", see also Remue, Hughes, De Houwer, & De Raedt, 2014; van der Kaap-Deeder, De Houwer, Hughes, Spruyt, & Vansteenkiste, 2018). In a similar study, Timko, England, Herbert, and Forman (2010) conducted two experiments that employed the IRAP to investigate general self-esteem and self-related attitudes toward body image. Again, both experiments comprised label stimuli that specified responding to self in terms of the phrases "I am" and "I am not." In other research, the participant's name or the name of a specified other has been employed to cue responding to self versus others, respectively. For example, Scanlon, McEnteggart, Barnes-Holmes, and Barnes-Holmes (2014) assessed gender bias and self-esteem in children, using an IRAP with label stimuli comprising the child's own name and a common name of the opposite gender, along with positive or negative target stimuli. And, Vahey, Barnes-Holmes, Barnes-Holmes, and Stewart (2009) employed a self-esteem IRAP that required participants to choose their own name (e.g., "David") or not their own name (e.g., "Not David") in relation

to either positive or negative target words. In general, the studies that have used the IRAP to examine the distinction between self versus others have found differences in responding to self and others, and these effects have correlated to some extent with other relevant indicators of self-esteem (see Vahey, Nicholson, & Barnes-Holmes, 2015 for a recent meta-analysis of the validity of the IRAP in the clinical domain).

In one of the most recent studies of self versus other using the IRAP, Barbero, López, Luciano, and Eisenbeck (2016) presented participants with their own names and the name of the researcher as label stimuli, and statements pertaining to specific characteristics of the self versus other as targets (e.g., "is in front of the laptop"). There were two response options ("Yes" and "No") on each trial. The four trial types in this study were referred to as: *I-I* (participant name-participant characteristics), *Other-Other* (researcher name-researcher characteristics), *I-Other* (participant name-researcher characteristics), and *Other-I* (researcher name-participant characteristics). The IRAP effects for all four trial types were broadly consistent with the actual experimental context. For example, participants responded more quickly when they had to respond that the laptop presenting the IRAP was in front of them, and not in front of the researcher.

In a systematic replication of the Barbero *et alii* (2016) study, Kavanagh, Barnes-Holmes, Barnes-Holmes, McEnteggart, and Finn (2018) used a similar IRAP, as well as a control IRAP that did *not* require responding to self versus other. That is, instead of comparing self with other, the control IRAP compared responding to two separate others (i.e., the researcher and a picture of another unknown participant). In Experiment 1, the data from the IRAP showed effects for both self and other that were again consistent with the experimental context, but only for two trial types (*I-I* and *Other-Other*), rather than all four trial types. A similar pattern was observed for the control IRAP (i.e., two of the trial type effects were significantly different from zero and two were not). In a second experiment, Kavanagh *et alii* (2018) produced results that more closely resembled those reported in the original study by Barbero *et alii* (2016) (i.e., significant effects for all four trial types).

In pursuing IRAP research that has focused on assessing responses to self versus other, an issue arose concerning the nature of the stimuli we were using, which we hoped to begin to address in the current work. Specifically, the present study was deemed to be exploratory in seeking to determine if it was feasible to use matched pictures of self versus others across two IRAPs. Specifically, we sought to determine if the IRAPs would produce intuitively predictable effects when pictures of self and matched pictures of others were inserted into the procedure (see below). The broader motivation for the study stemmed from an on-going research program that seeks to develop the IRAP as a measure of perspective-taking (Kavanagh *et alii*, 2018; Kavanagh *et alii*, in press; see also Barbero *et alii*, 2016). In each of these studies, the stimuli that were presented in the IRAP all consisted of words or statements that pertained to the characteristics of the participant or another individual. Typically, self-related terms involved using the participant's name or words such as "I", "my" or "me", whilst other-related terms involved using another's name or words such as "they", "others" or "them." Although this research yielded interesting findings, we had concerns that the use of such stimuli might allow for some ambiguity in how these stimuli were interpreted by participants. For example, when the on-screen stimulus was "I" the assumption was that the participant would interpret this as referring to self, rather than to the computer or another person. In general, it appeared that this assumption was upheld, but of course room for ambiguity remained. One way in which we thought we could remove this potential ambiguity was

to insert a picture of the participant into the IRAP itself, so that on those trials when the picture of the participant appeared, it would be clearly interpreted as a self-related stimulus for the participant. If a participant's picture was shown to possess at least some of the functional properties of self, then the potential ambiguity arising from previous studies that used self-referential terms could be removed or at least reduced.

In the current study, participants were asked to bring an electronic picture of their face which could be inserted into the IRAP at the beginning of the session. They were also asked to bring a picture of another person, who they considered to be relatively similar to themselves in terms of age, gender, and general facial features (i.e., a 'matched' picture). Each participant was then exposed to two IRAPs, with one IRAP containing the picture of themselves, and the other IRAP containing the matched picture of the other person. Both IRAPs presented the same pictures of pens as the contrast category. The two IRAPs were thus very similar in that they both involved confirming and disconfirming if the on-screen pictures contained faces versus pens. We expected to find IRAP effects that were consistent with previously-established verbal relations. That is, participants would confirm more readily than deny that a picture of a face was a face, and that a picture of a pen was a pen, as well as denying more readily than confirming that a picture of a pen was a face and that a picture of a face was a pen. Given the exploratory nature of the study, we made no formal predictions concerning the extent to which the two IRAPs (self-picture versus other-picture) would produce different outcomes. Although the results were broadly consistent with common-sense expectations, specific patterns emerged that connect with very recent conceptual issues surrounding the use and development of the IRAP as a measure of relational responding per se. On the grounds of intellectual honesty, we will address these issues in the Discussion, rather than in the Introduction.

METHOD

Participants

Thirty-four participants were recruited for the current experiment, 29 females and 5 males. Participants ranged from 17-45 years (M= 21.4) and were recruited through random convenience sampling from the Ghent University participant pool. Each participant was paid an hourly rate of 10 euro. Given that this was the first study in which pictures of the self versus others had been inserted into an IRAP, it was not possible to conduct a meaningful power analysis. Nevertheless, the general strategy for recruiting numbers of participants was guided by the results of a recent meta-analysis of IRAP effects in the clinical domain, indicating that a minimum of 29 participants is required to achieve a power of 0.8 for first-order correlations (Vahey *et alii*, 2015). Because participants sometimes failed to reach various performance criteria for the IRAP (details provided subsequently), it was necessary to recruit more than 29 participants to yield an adequate dataset for analyses.

Materials and Apparatus

The experiment comprised two IRAPs, a self-picture IRAP and an other-picture IRAP. Six self-report questionnaires pertaining to a range of clinical and self/other constructs were also used, in a purely exploratory manner (i.e., because they had been

used in other studies that formed part of a larger research program). Hence, no specific hypotheses regarding these measures were made and details of these measures are not provided in the text. As an aside, only one significant correlation (out of a total of 50) at the .05 level emerged between the IRAP trial types and one self-report measure, and thus these correlational analyses are not provided in this article (further details available from the first author upon request).

Picture stimuli used in the IRAPs. The face picture stimuli were collected for both IRAPs before the experiment began. Participants were asked to bring to the experiment two pictures; a picture of themselves that they liked and a picture of an unknown other who they considered to be similar in looks to themselves (i.e., same gender, age, hair color, skin color, and eye color). These pictures were included in the self-picture and other-picture IRAPs, respectively. No formal measures were taken of the similarity between the pictures of self and other, as chosen by the participant, but the pictures were checked by the Researcher beforehand to determine that the pictures were broadly similar. In one case, a male participant with a beard and short hair selected a 'matched' picture of a female with long hair; for ethical reasons this participant completed the study but the data were not included in any of the analyses.

Self-picture IRAP. The self-picture IRAP presented a total of six label stimuli on the top of the screen; three of the labels were words pertaining to faces (i.e., "face", "head", and "person"), with the remaining three words pertaining to pens (i.e., "bic", "pen", and "stylo", see Table 1). The target stimuli consisted of four images presented in the center of the screen, one picture depicted the individual participant's face and the other three pictures depicted three different pens. The response options "Yes" and "No" were presented at the bottom left- and right-hand corners. The four trial types were denoted as: *Face words-Self picture*, *Face words-Pen pictures*, *Pen words-Self picture*, and *Pen words-Pen pictures* (see Figure 1).

Table 1. Label stimuli for the self-picture and other-picture IRAPs.

Face words	Pen words
Person	Pen
Face	Stylo
Head	Bic

Note: Stimuli were presented to participants in Dutch.

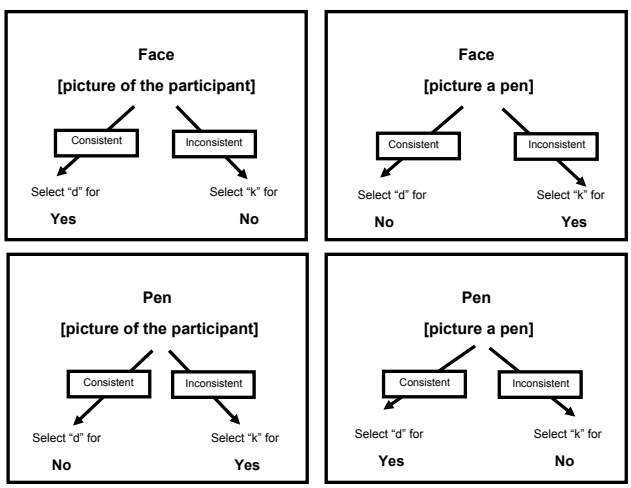

Figure 1. Examples of the four trial types in the self-picture IRAP: *Face words-Face picture*, *Face words-Pen pictures*, *Pen words-Face picture*, and *Pen words-Pen pictures*. The words *Consistent* and *Inconsistent* were not shown on-screen.

Other-picture IRAP. The other-picture IRAP was similar to the self-picture IRAP, except that the four target pictures now depicted the face of another person (rather than the face of each participant), as well as the three different pens. The four trial types were denoted as: *Face words-Other picture*, *Face words-Pen pictures*, *Pen words-Other picture*, and *Pen words-Pen pictures* (see Figure 2).

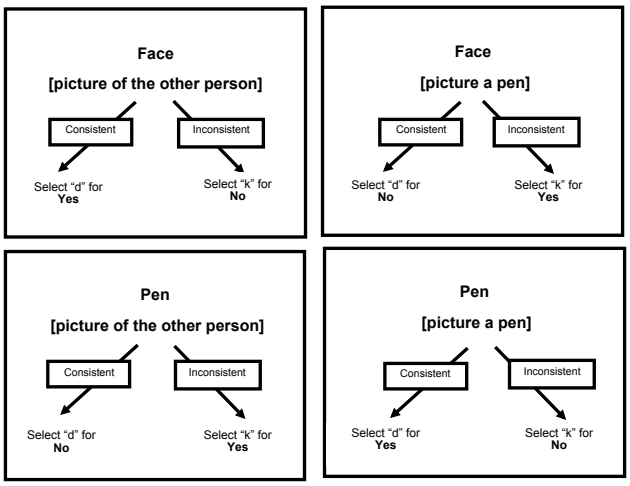

Figure 2. Examples of the four trial types in the other-picture IRAP: *Face words-Face picture*, *Face words-Pen pictures*, *Pen words-Face picture*, and *Pen words-Pen pictures*. The words *Consistent* and *Inconsistent* were not shown on-screen.

Procedure

The experiment took place on an individual basis in sound-proof cubicles at Ghent University. Informed consent was obtained from all participants. Before presenting the IRAPs, the picture stimuli were uploaded to the IRAP program. Each participant was exposed to the self-picture and the other-picture IRAPs, with the order of each counterbalanced across participants. Participants thereafter completed the battery of questionnaires.

Self-picture IRAP. The self-picture IRAP comprised a maximum of four pairs of practice blocks, followed by three pairs of test blocks. The following instructions were presented at the start of the IRAP program: "This task will present sets of words or images. You will be asked to relate the words or images. If you make a mistake, you'll see a red X. Provide the correct response to continue. Respond as accurately as you can. When you've learned to be accurate you'll naturally speed up too." On each trial, the label (face-related word or pen-related word) appeared at the top of the screen, with a target (picture of participant's face or picture of a pen) in the center, and the two response options ("Yes" and "No") at the bottom. Participants responded on each trial using either the "d" key for the response option on the left or the "k" key for the response option on the right. The locations of the response options alternated from trial to trial in a quasi-random order, such that they did not remain in the same left-right locations for more than three successive trials. The instruction "The previously correct and incorrect answers have been reversed" was presented between blocks of trials. The instruction "Continue responding both as accurately and as quickly as you can" was presented between block pairs of trials.

When participants selected the response option that was deemed correct within that

block, an inter-trial interval of 400 ms was presented, after which the next trial occurred. When participants selected the response option that was deemed incorrect for that block, the stimuli remained on the screen and a red "X" appeared beneath the target stimulus. Only when the correct response option was selected did the program proceed to the 400 ms inter-trial interval (followed by the next trial). This pattern of trial presentations, with corrective feedback, continued until the entire block of 32 trials was presented. Trials were presented in a quasi-random order within each block, such that each of the four trial types appeared eight times within each block. Consistent blocks required responding that was in accordance with pre-experimental verbal relations: *Face words-Face picture/Yes*, *Face words-Pen pictures/No*, *Pen words-Face picture/No*, and *Pen words-Pen pictures/Yes*. Inconsistent blocks required the opposite: *Face words-Face picture/No*, *Face words-Pen pictures/Yes*, *Pen words-Face picture/Yes*, and *Pen words-Pen pictures/No*. The presentation of consistent and inconsistent blocks was counterbalanced across participants.

When participants completed a block of trials, the IRAP program delivered feedback on their performance during that block. The feedback consisted of a message informing them how accurately and how quickly they had responded. The latter was calculated from stimulus onset to the first correct response across all 32 trials within the block. Participants were required to achieve a minimum accuracy of 80% correct and a maximum median latency of no more than 2000 ms on each block. If participants achieved both accuracy and latency criteria on the first, second, third, or fourth pair of practice blocks, they proceeded to the first pair of test blocks; if they failed on the fourth pair of practice blocks participation in the experiment was terminated.

A fixed set of six test blocks was presented with no accuracy or latency criteria required for participants to progress from one block to the next. However, percentage correct and median latency were presented at the end of each block to encourage participants to maintain the accuracy and latency levels they had achieved during the practice blocks.

Other-picture IRAP. The other-picture IRAP was similar to the self-picture IRAP, except that it contained the picture of the other face rather than the face of the participant.

Questionnaires. Participants completed the six questionnaires immediately after completing the two IRAPs.

All procedures in the current study were in accordance with the ethical standards of the institutional research committee, and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Results

A summary of the means and standard deviations for the six questionnaires is provided in Table 2. The mean score on the Self-warmth Thermometer was around the mid-way point at 56.72 (/100) indicating relatively average warmth toward self. The overall Psychological Flexibility Index (PFI) scores were relatively high, indicating high psychological flexibility. The overall Community Assessment of Psychic Experiences (CAPE) and subscale scores were relatively low, indicating low psychotic-like symptoms. The Experiences in Close Relationships-Relationship Structures questionnaire (ECR-RS) scores were also relatively low in terms of both attachment-anxiety and attachment-avoidance. The overall Experiencing of Self Scale (EOSS) and subscale scores were low, indicating low control by others over the experience of self. The Inclusion of Other in the Self (IOS) scores for best friend were higher than for other people, suggesting a closer relationship in this regard. Nothing unusual or unexpected, therefore, emerged from the questionnaires, given the use of a non-clinical sample.

Consistent with standard practice in IRAP research, mean response latencies for consistent and inconsistent blocks were initially divided according to trial type and

Table 2. Descriptive statistics for questionnaires.

Questionnaire		Direct scores	*D*
Self-warmth Thermometer		56.72	22.82
Psychological Flexibility Index (PFI)		353.59	27.15
Community Assessment of Psychic Experiences (CAPE, weighted scores)	Overall Frequency	1.76	.35
	Frequency of Positive Symptoms	1.49	.42
	Frequency of Negative Symptoms	1.93	.39
	Frequency of Depressive Symptoms	2.14	.51
	Overall Distress	2.26	.51
	Distress associated with Positive Symptoms	1.8	.54
	Distress associated with Negative Symptoms	2.15	.58
	Distress associated with Depressive Symptoms	2.73	.68
Experiences in Close Relationships, Relationship Structures Questionnaire (ECR-RS)	Attachment-related avoidance (Mother)	17.94	9.15
	Attachment-related anxiety (Mother)	5.313	3.92
	Attachment-related avoidance (Father)	21.84	9.61
	Attachment-related anxiety (Father)	6.38	4.26
	Attachment-related avoidance (Partner)	10.34	4.48
	Attachment-related anxiety (Partner)	8.53	4.64
	Attachment-related avoidance (Best Friend)	12.13	5.72
	Attachment-related anxiety (Best Friend)	6.34	3.34
Experiencing of Self Scale (EOSS)	Overall EOSS	74.63	15.08
	Casual acquaintances-absent	18.13	5.64
	Casual acquaintances-present	24.25	4.26
	Close relationships-absent	11.19	6.08
	Close relationships – present	21.06	4.59
Inclusion of Other in the Self scale (IOS)	Best friend	5.06	1.31
	Other people	3.22	1.34

Notes: The maximum score is 100 for the Self-warmth Thermometer; The maximum score for the PFI is 492. The maximum weighted score for all CAPE subscales is 4.00; The maximum score for each of the ECR-RS attachment related avoidance subscales is 42 and for the Attachment related anxiety subscale the maximum is 21; The maximum overall EOSS score is 140 with the maximum score for each subscale at 35; The maximum score for each of the IOS scales is 7; None of the scales have formal clinical cut-offs.

calculated for each participant (see Barnes-Holmes, Barnes-Holmes, Stewart, & Boles, 2010). Based on the latency and accuracy criteria, one participant failed the practice blocks on the self-picture IRAP and was excluded from further analysis. Of the remaining participants, one failed to maintain criteria across the test block pairs in the self-picture IRAP, again these data were excluded from further analysis (final N= 32).

Consistent with the majority of published IRAP studies, D_{IRAP}-scores for both IRAPs were calculated for each of the four trial types (see Barnes-Holmes *et alii*, 2010), such that positive D_{IRAP}-scores during consistent blocks indicated responding "Yes" more quickly than "No" on *Face words-Face picture* and *Pen words-Pen pictures* trial types, and responding "No" more quickly than "Yes" on *Face words-Pen pictures* and *Pen words-Face picture* trial types. Negative D_{IRAP}-scores indicated the opposite pattern: responding "No" more quickly than "Yes" on *Face words-Face picture* and *Pen words-Pen pictures* trial types, and responding "Yes" more quickly than "No" on *Face words-Pen pictures* and *Pen words-Face picture* trial types.

The mean D_{IRAP}-scores and standard errors for each trial type are presented in Figure 3. Positive scores were recorded for all four trial types, with the weakest on *Pen words-Face picture* in both of the IRAPs. For each of the four trial types, the difference between the two IRAPs appeared relatively modest. A 2x4 mixed repeated measures ANOVA produced a main effect for trial type [$F(1, 31)= 17.14$, $p <.001$, $\eta_p^2= .36$], but not for IRAP type ($p >.05$), or for the interaction ($p >0.2$). Post-hoc comparisons, with the trial type effects collapsed across the self- and other-picture IRAPs, indicated that each trial type differed from every other trial type, except for the comparison between

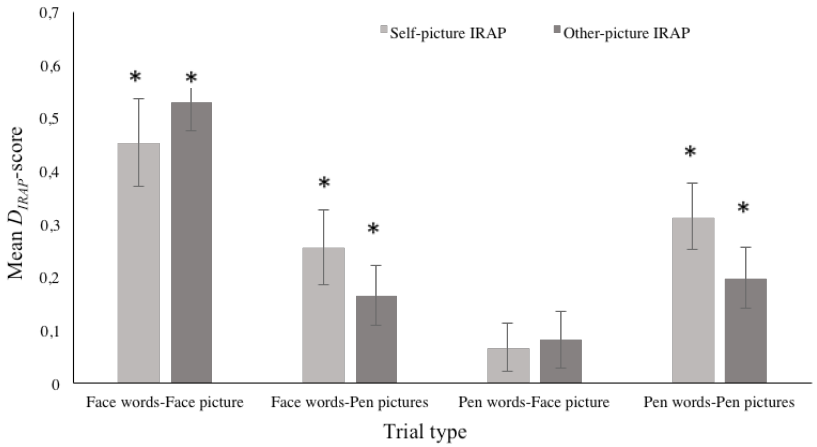

Figure 3. Mean *DIRAP*-scores on the self-picture and the other-picture IRAP trial types. Positive *DIRAP*-scores indicate history-consistent responding. * Indicates *DIRAP*-scores that are significantly different from zero.

Faces words-Pen pictures and *Pen words-Pen pictures*. Eight planned one-sample *t*-tests indicated that the effects were significantly different from zero (*p*s <.05) for both IRAPs for three of the four trial types; the effects were non-significant for the *Pen words-Face pictures* trial type.

Given that there was no significant main effect for IRAP type or interaction with trial type, a single overall *D*-score (the mean of the four trial types) was calculated for each IRAP, and then subjected to correlational analyses with the questionnaires (a total of twenty-five correlations for each IRAP). Neither set of correlations proved to be significant (all *p*s >.05).

DISCUSSION

The primary objective of the current study was to determine if it was feasible to use matched pictures of self versus other across two IRAPs. In both IRAPs, the contrast categories were pen-related stimuli. In general, the pattern of effects indicated that pictures of faces could be incorporated into the IRAP, thus supporting the feasibility of using photographs of self versus other in IRAP perspective-taking studies. As noted briefly before, only one significant correlation (out of a total of 50) emerged between the IRAP trial types and the self-report measures, and thus it should be interpreted with extreme caution and seems unworthy of further discussion. However, two effects did emerge that are worthy of further consideration. Before discussing these effects, it seems important to speculate as to why no difference emerged between the self- and other-IRAPs (even though such differences were not predicted).

In previous research, we and others have found differences between reactions to self versus others, using IRAPs (Barbero, *et alii*, 2016; Kavanagh, *et alii*, 2018). In both cases, however, the research involved comparing responses to self versus others *within* an IRAP, rather than across two separate IRAPs, as was the case in the current study. It is possible, therefore, that the context for discriminating between self and others is much reduced when the current strategy is adopted. More informally, in the current research participants' responding in the IRAPs was controlled more by the presence of a face

versus a pen, rather than by the presence of one's own face versus the face of another as would be the case in an IRAP that required responding to the participant's own face versus another person's face. In other words, it would be interesting to determine if differential trial type effects were observed when pictures of self versus others were presented within the same IRAP. For example, it is possible that a larger IRAP effect might be observed for a trial type that presented a self-related term with a picture of the participant than a trial type that presented an other-related term with a picture of a different individual (see below for a discussion of such differential trial type effects).

Leaving aside the lack of difference between self and other in the current study, as noted above, two effects did emerge that are interesting in their own right, one that has already been reported in the literature, and one that has yet to be articulated. Before considering each of these effects, we will summarize the overall pattern of trial type effects observed for both self- and other-IRAPs. Specifically, with the exception of one trial type, the effects recorded were consistent with what might be expected based on previously-established verbal relations. That is, participants tended to confirm more readily than to deny that a picture of a face was a face, and that a picture of a pen was a pen. They also tended to deny more readily than confirm that a picture of a pen was a face. However on both IRAPs, participants failed, at a statistically significant level, to deny more quickly than confirm that a pen-related word was a picture of a face. Post-hoc tests that compared the two trial types that required participants to deny a relation more readily than to confirm it proved to be statistically significant, highlighting a potentially important difference that needs to be explained. We will return to this issue subsequently.

The first effect we will consider has been referred to as the Single-Trial-Type-Dominance-Effect (STTDE) and was observed in the current study (see also Finn, Barnes-Holmes, & McEnteggart, 2018, Kavanagh *et alii*, 2018, Kavanagh *et alii*, in pres). Specifically, the *D*-score for the *Face words-Face picture* trial type was significantly larger than for the other three trial types. The fact that the *Face words-Face picture* effect was larger than both *Face words-Pen picture* and *Pen words-Face picture* may be deemed relatively predictable because the latter two trial types do not share the same response option with the former trial type within blocks of trials. During consistent bocks, for example, the *Face words-Face picture* trial type requires the participant to respond Yes, whereas the *Face words-Pen picture* and *Pen words-Face picture* trial types require the response No. What is more difficult to explain, however, is the difference between the *Face words-Face picture* and the *Pen words-Pen picture*, because they both share the same response option within IRAP blocks (i.e., "Yes" during consistent blocks and "No" during inconsistent blocks).

In a previously published study (Kavanagh *et alii*, 2018), we explained this effect using a recently-proposed model of differential IRAP effects that are sometimes observed. The model is known as the Differential Arbitrarily Applicable Relational Responding Effects (DAARRE) model and it could be used to explain the STTDE observed in the current study (see Figure 4). The explanation is based on the assumption that in general participants oriented toward face-related stimuli more strongly than toward pen-related stimuli (see Borji, Sihite, & Itti, 2013; Hershler, & Hochstein, 2005; Santos, Mier, Kirsh, & Meyer-Lindenberg, 2011). We use the term 'face-related' stimuli based on the assumption that words coordinated with pictures of faces in the context of the IRAP would possess some of the functions (e.g., orienting) of those pictures, via a transfer of functions. The term 'oriented' is used here simply to suggest that in general human faces would be more salient or 'attention-grabbing' than pens. In other words, face-related

stimuli could be seen as having stronger orienting functions relative to pen-related stimuli.

The model identifies three key sources of behavioral influence: (1) the relationship between the label and target stimuli (labeled as Crels), (2) the orienting functions of the label and target stimuli (labeled as Cfuncs), and (3) the coherence functions of the two response options (e.g., "Yes" and "No"). Consistent with the assumption that face stimuli likely possess stronger orienting functions relative to pen stimuli, the Cfunc property for face is labeled as positive and the Cfunc property for pens is labeled as negative. The negative labeling for pens should not be taken to indicate a negative orienting function, but simply an orienting function that is weaker than that of faces. The labeling of the relations between the label and target stimuli indicates the extent to which they cohere or do not cohere based on the participants' relevant verbal history. Thus, a *face-face* relation is labeled with a plus sign (i.e., coherence), whereas a *face-pen* relation is labeled with a minus sign (i.e., incoherence). Finally, the two response options are each labeled with a plus or minus sign to indicate their functions as either coherence or incoherence indicators (see Maloney & Barnes-Holmes, 2016). In the current example, "Yes" (+) would typically be used in natural language to indicate coherence, and "No" (-) would be used to indicate incoherence. Note, however, that these and all of the other functions labeled in Figure 4 are behaviorally determined, by the past and current verbal history of the participant, and should not be seen as absolute or inherent in the stimuli themselves.

As can be seen from Figure 4, each trial type differs in its pattern of Cfuncs and Crels, in terms of plus and minus properties, that define the trial type for the IRAPs. The STTDE for the *Face words-Face picture* trial type may be explained, as noted above, by the DAARRE model based on the extent to which the Cfunc and Crel properties cohere with the relational coherence indicator (RCI) properties of the response options across blocks of trials. To appreciate this explanation, note that the Cfunc and Crel properties for the *Face words-Face picture* trial type are all labeled with plus signs;

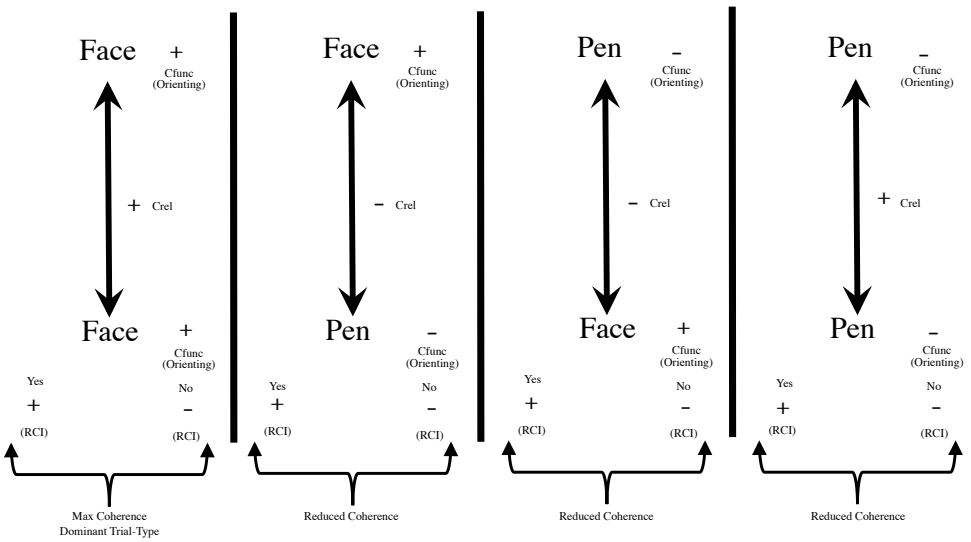

Figure 4. The DAARRE model as it applies to both the self-picture and other-picture IRAPs. The positive and negative labels refer to the relative positivity of the Cfuncs, for each label and target, the relative positivity of the Crels, and the relative positivity of the RCIs in the context of the other Cfuncs, Crels, and RCIs in that stimulus set.

in addition, the RCI that is deemed correct for history-consistent trials is also labeled with a plus sign (the only instance of four plus signs in the diagram). In this case, therefore, according to the model this trial type may be considered maximally coherent during history-consistent trials. In contrast, during history-inconsistent trials, there is no coherence between the required RCI (minus sign) and the properties of the Cfuncs and Crel (all plus signs). According to the DAARRE model, this stark contrast in levels of coherence across blocks of trials serves to produce a relatively large IRAP effect. Now consider the *Pen words-Pen pictures* trial type, which requires that participants choose the same RCI as the *Face words-Face picture* trial type during history-consistent trials, but here the property of the RCI (plus sign) does not cohere with the Cfunc properties of the label and target stimuli (both minus signs). During history-inconsistent trials, the RCI (minus sign) does cohere with the Cfunc properties, but not with the Crel property (plus sign). Thus, the differences in coherence between history-consistent and history-inconsistent trials across these two trial types is not equal (i.e., the difference is greater for the *Face words-Face picture* trial type) and thus favors the STTDE (for *Face words-Face picture*). Finally, as becomes apparent from inspecting Figure 4 for the remaining two trial types (*Face words-Pen pictures* and *Pen words-Face picture*), the differences in coherence across history-consistent and history-inconsistent blocks is reduced relative to the *Face words-Face picture trial* type (two plus signs relative to four), thus again supporting the STTDE.

At this point, we have used the DAARRE model to explain the STTDE that emerged once again in the current study, but as noted above, another differential trial type effect also emerged that would be difficult to explain in terms of the response options alone. Specifically, participants tended to deny more readily than confirm that a picture of a pen was a face, but this was not the case (at a statistically significant level) when denying that a picture of a face was a pen. Indeed, post-hoc analyses indicated that these two trial types differed significantly from each other. How might we explain this difference, given that both trial types required the same RCI within blocks of trials? Once again, the DAARRE model may be of use here. If we examine Figure 4, it becomes apparent that the *Face words-Pen pictures* trial type presents a target stimulus that coheres with the RCI in terms of its Cfunc properties, whereas the *Pen words-Face picture* trial type does not. If we assume that the spatial contiguity between the target stimulus and the response option plays a role in determining the IRAP effect, the difference in trial type effects observed here makes sense. More informally, participants may experience a "*Yes-No-No*" reaction to the *Face words-Pen pictures* trial type, but a "*No-Yes-No*" reaction to the *Pen words-Face picture trial* type, assuming that in general they read each IRAP trial from top-down. If participants find it easier to select an RCI that is functionally similar to the target stimulus they have just observed, than an RCI that is functionally *dissimilar*, the larger effect for the *Face words-Pen pictures* trial type is readily predicted. We refer to this effect as the Dissonant Target Trial Type Effect (DTTTE, Finn & Barnes-Holmes, 2019).

Of course, the DAARRE model explanation presented above is post-hoc and speculative, but it seems important to present it here because it is a pattern we have observed in other studies, in which the DAARRE model can be applied in a relatively straightforward way. Perhaps other researchers, therefore, who are using the IRAP, may find the interpretation offered here of some use in attempting to explain and explore similar effects. In any case, it seems important to continue to develop increasingly sophisticated functional analyses of the IRAP in terms of the cluster of variables that

produce the patterns we observe with the measure. Indeed, this seems particularly important because the IRAP has been used widely as a measure in clinical, health, forensic, and social psychological research (e.g., see Vahey *et alii*, 2015 for a recent meta-analysis).

References

Aron A, Aron EN, & Smollan D (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology*, *63*, 596-612. Doi: 10.1037/0022-3514.63.4.596

Barbero-Rubio A, López-López J, Luciano C, & Eisenbeck N (2016). Perspective-taking measured by implicit relational assessment procedure (IRAP). *The Psychological Record*, *66*, 1-10. Doi: 10.1007/s40732-016-0166-3

Barnes-Holmes D, Barnes-Holmes Y, Stewart I, & Boles S (2010). A sketch of the implicit relational assessment procedure (IRAP) and the relational elaboration and coherence (REC) model. *The Psychological Record*, *60*, 527-542. Doi: 10.1007/BF03395726

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. Doi: 10.1007/BF03395634

Borji A, Sihite DN, & Itti L (2013). What stands out in a scene? A study of human explicit saliency judgment. *Vision Research*, *91*, 62-77. Doi: 10.1016/j.visres.2013.07.016. Epub 2013 Aug 15.

Finn M, & Barnes-Holmes D (2019). *Predicting-and-influencing patterns of arbitrarily applicable relational responding in individual performances in the Implicit Relational Assessment Procedure*. Paper presented at the Association for Contextual Behavioral World Conference, Dublin, Ireland.

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

Hershler O & Hochstein S (2005). At first sight: A high-level pop out effect for faces. *Vision Research, 45,* 1707-1724. Doi: 10.1016/j.visres.2004.12.021

Hughes S & Barnes-Holmes D (2016). Relational frame theory: The basic account. In R Zettle, SC Hayes, D Barnes-Holmes, & T Biglan (Eds.), *Wiley Handbook of Contextual Behavioral Science* (pp. 176-232). Cambridge, MA: Wiley-Blackwell. Doi: 10.1002/9781118489857.ch9

Hussey I, Barnes-Holmes D, & Barnes-Holmes Y (2015). 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

Kavanagh D, Barnes-Holmes Y, Barnes-Holmes D, McEnteggart C, & Finn M (2018). Exploring differential trial type effects and the impact of a read-aloud procedure on deictic relational responding on the IRAP. *The Psychological Record, 68*, 163-176. Doi: 10.1007/s40732-018-0276-1

Kavanagh D, Roelandt A, Van Raemdonck L, Barnes-Holmes Y, Barnes-Holmes D, & McEnteggart C (in press). The on-going search for perspective-taking IRAPs: Exploring the potential of the natural language IRAP. *The Psychological Record*.

Maloney E & Barnes-Holmes D (2016). Exploring the behavioral dynamics of the implicit relational assessment procedure: The role of relational contextual cues versus relational coherence indicators as response options. *The Psychological Record, 66*, 1-9. Doi: 10.1007/s40732-016-0180-5

Remue J, De Houwer J, Barnes-Holmes D, Vanderhasselt M, & De Raedt R (2013). Self-esteem revisited: Performance on the implicit relational assessment procedure as a measure of self-versus ideal self-related cognitions in dysphoria. *Cognition and Emotion, 27*, 1441-1449. Doi: 10.1080/02699931.2013.786681

Remue J, Hughes S, De Houwer J, & De Raedt R (2014). To be or want to be: Disentangling the role of actual versus ideal self in implicit self-esteem. *PLoS ONE, 9*, e108837. Doi: 10.1371/journal.pone.0108837

Santos A, Mier D, Kirsch P, & Meyer-Lindenberg A (2011). Evidence for a general face salience signal in human amygdala. *Neuroimage*, *54*, 3111-3116. Doi:10.1016/j.neuroimage.2010.11.024

Scanlon G, McEnteggart C, Barnes-Holmes Y, & Barnes-Holmes D. (2014). Using the implicit relational assessment procedure (IRAP) to assess implicit gender bias and self-esteem in typically-developing children and children in ADHD and with dyslexia. *Behavioral Development Bulletin, 19*, *48-59*. Doi: 10.1037/h0100577

Timko CA, England EL, Herbert JD, & Forman EM (2010). The implicit relational assessment procedure as a measure of self-esteem. *The Psychological Record, 60,* 679-698. Doi: org/10.1007/BF03395739

Vahey N, 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. Doi: org/10.1007/BF03395670

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

Van der Kaap-Deeder J, De Houwer J, Hughes S, Spruyt A, & Vansteenkiste M (2018). The development and validation of an implicit measure of competence need satisfaction. *Motivation and Emotion*, *42,* 615-637. Doi: 10.1007/s11031-018-9685-3

World Health Organisation (2017). Process of translation and adaptation of instruments. Retrieved from http://www.who.int/substance_abuse/research_tools/translation/en/

Received, March 18, 2019
Final Acceptance, July 29, 2019