



# Do Response Options in the Implicit Relational Assessment Procedure (IRAP) Matter? A Comparison of Contextual Relations versus Relational Coherent Indicators

Emma Maloney<sup>1</sup> · Mairéad Foody<sup>2</sup> · Carol Murphy<sup>3</sup>

Published online: 25 November 2019

© Association for Behavior Analysis International 2019

## Abstract

Empirical analysis of features of the Implicit Relational Assessment Procedure may be important to results. As such, the current research compared effects of response options that were contextually cued relational responses ( $C_{relS}$ ) versus relational coherence indicators (RCIs) across two IRAPs conducted with college student participants ( $N = 40$ ). The IRAPs were similar except for the response options used, which were either “Same”/“Opposite” ( $C_{relS}$ ) versus “Accurate”/“Inaccurate” (RCIs).  $D$ -scores for both IRAPs showed the expected IRAP effect (bias). A critical difference was noted dependent upon the type of response options used: the IRAP effect was shown to be stronger when  $C_{rel}$  response options were used. There was no statistically significant interaction effect shown between response option used and order of completion (i.e.,  $C_{rel}$  IRAP first vs. RCI IRAP first), however, there was a statistically significant interaction effect shown between type of response options used on the IRAP, order of completion, and block-order presentation (consistent trial-blocks vs. inconsistent trial blocks presented first). Findings are discussed regarding potential implications and further research.

**Keywords** IRAP ·  $C_{relS}$  · RCIs · response options

The Implicit Relational Assessment Procedure (IRAP; Barnes-Holmes et al. 2006) is the first behavior-analytic measure to contribute to the research literature in implicit cognition across a wide range of socially sensitive topics. IRAP research has examined implicit bias towards homo- and heterosexuality (Cullen and Barnes-Holmes 2008), implicit beauty bias (Murphy et al. 2014), implicit racial bias (Barnes-Holmes et al. 2010), implicit self-esteem (Remue et al. 2013), and implicit bias toward sexualization of children among sexual offenders (Dawson et al. 2009). A sizeable amount of IRAP research has been directed towards refining and thus strengthening the IRAP as a measure of implicit bias. Research has shown that the IRAP is difficult to fake

(McKenna et al. 2007), is comparable to the Implicit Association Test (IAT; Greenwald et al. 1998) as a measure of individual differences (Barnes-Holmes et al. 2009), and is a valid measure of implicit cognition (Golijani-Moghaddam et al. 2013).

An examination of the components of the IRAP and how they function (Campbell et al. 2011) systematically manipulated stimulus presentations in the IRAP and found that the IRAP effect varied in strength dependent on the positioning of sample stimuli, target stimuli, and response options (e.g., fixed or alternating position). Further, Barnes-Holmes et al. (2010) determined that the time constraints imposed in the IRAP had an influence on participant responding. Implicit measures frequently involve participants responding under time pressure and what is often referred to as “automatic” responding, which may differ from more deliberative responding. For example, an antirace bias may be evident in participant responding on an implicit measure, but not in their self-report data. Such findings may be due in part to the self-report methodology, which allows for more deliberation and consideration. The Relational Elaboration and Coherence model (Barnes-Holmes et al. 2010), which laid the foundation for the IRAP, attempts to account for “automatic” responding in terms of

✉ Carol Murphy  
carol.a.murphy@mu.ie

Mairéad Foody  
mairead.foody@dcu.ie

<sup>1</sup> National University of Ireland, Maynooth, Ireland

<sup>2</sup> Dublin City University, Dublin City, Ireland

<sup>3</sup> National University of Ireland, Maynooth, Ireland

brief immediate relational responding (BIRR), as required in IRAP tasks, compared to extended elaborated relational responding (EERR), as when an individual has sufficient time to modify their BIRRs. For example, the model suggests that when completing an implicit measure under time pressure stereotype responding may be evident (e.g., participants' BIRRs show more favorable bias affirming "men/science" compared to "women/science"), but similar bias may not be evident in their self-report responses partly because the absence of time pressure allows for additional, more sophisticated, prelearned relations (EERRs) to come into play (e.g., "prejudging scientific ability based on sex/unwise/unjust").

The impact of the use of types of response options has also been examined (Maloney and Barnes-Holmes 2016), and the expected implicit bias was shown to be stronger under conditions with response options that were contextually related stimuli ( $C_{rel}$ s such as Similar/Opposite) versus relational coherence indicators (RCIs such as True/False). The IRAP task requires participants to alternate between affirming stereotypes (e.g., thin-positive-similar/fat-positive-opposite) and anti-stereotypes (e.g., thin-negative-similar/ fat-negative-opposite) when trials illustrating relations of this type are presented on a computer-screen. Early IRAP research used  $C_{rel}$ s, (e.g., same-opposite); however, recent IRAP researchers have tended to use RCIs (e.g., True/False; Right/Wrong; Yes/No; see Hussey et al. 2015; McEnteggart et al. 2016; Hussey and Barnes-Holmes 2012), thus it is important to examine whether  $C_{rel}$ s/RCIs are interchangeable with no effect on the research, or if there is a potential impact on IRAP findings depending on which type of response options are used.

To understand the significance of this research it is important to briefly discuss the theoretical framework from which the IRAP was developed; relational frame theory (RFT; Hayes et al. 2001). RFT drew from earlier work on stimulus equivalence (see Sidman 1971) to propose that, with the aid of relational cues, it is possible for humans to construct diverse patterns of relational responding (or relational frames) which serve as the basis for human language and cognition. According to RFT there are two types of contextual cues. Functional cues (or  $C_{func}$ s) are used to define the form of psychological properties that can be transformed across stimuli participating in a relational frame. For example, the recollection of what an orange tastes like, which is evoked in the presence of the verbal stimulus "Orange," is controlled by  $C_{func}$ s and is based upon the equivalence relation between the word "Orange" and the object "Orange," which evokes similar sensory stimulation. Examples of  $C_{func}$ s include "tastes like" or "looks like." The other form of contextual cues referred to in RFT are the above-mentioned  $C_{rel}$ s.  $C_{rel}$ s denote the type of relational response that is likely to be reinforced by the verbal community (e.g., "equal to" in the case of an equivalence relation). Words such as "same,"

"opposite," or phrases such as "bigger than," "smaller than" function as  $C_{rel}$ s because they directly imply the relation that exists between stimuli.

RFT literature distinguishes between contextual cues and terms such as "True" and "False" because the latter are considered to be evaluative terms, indicative of relational coherence (Hayes and Barnes 1997; Barnes-Holmes, Hayes, Dymond, & O'Hora, 2001). That is, a statement is considered to be true if it coheres with wider patterns of relational responding in natural language within the verbal community. In contrast, statements that do not cohere with expected language practices are considered to be false. It could therefore be surmised that the terms "True" and "False" (and potentially their variants) have a different verbal function compared to relational terms such as "Same" and "Opposite," in that the former operate as relational coherence indicators (RCIs) in some contexts (Maloney and Barnes-Holmes 2016) and not as  $C_{rel}$ s per se.

Although not a novel theoretical concept, this distinction had not been addressed in IRAP research until recently. Maloney and Barnes-Holmes (2016) compared the effects on two IRAPs with college students ( $N = 52$ ), one IRAP used  $C_{rel}$  response options (Similar/Different) and one used RCI response options (True/False). Both IRAPs were otherwise similar; the  $C_{rel}$  IRAP presented four relations (IRAP trial-types) using word stimuli as follows: Pleasant/Positive/Similar; Unpleasant/Positive/Different, Pleasant/Negative/Different, Unpleasant/Negative/Similar whereas the RCI IRAP presented Pleasant/Positive/True, Unpleasant/Positive/False, Pleasant/Negative/False, Unpleasant/Negative/True. It was expected that IRAP response latency data would show that participants more rapidly affirmed Pleasant-Positive and Unpleasant-Negative relations compared to the converse. The order in which participants completed both IRAPs was counterbalanced ( $n = 26$  were randomly assigned to complete the  $C_{rel}$  IRAP first, whereas  $n = 26$  completed the RCI IRAP first). The order of trial-blocks presented was also manipulated as is standard in IRAP research (e.g., Nicholson and Barnes-Holmes 2012). Maloney and Barnes-Holmes (2016) found no main effect reported for the type of response option used, however, the results indicated a significant interaction effect related to the order in which participants completed the two types of IRAP. In particular, the IRAP effect (expected bias) was reduced in the RCI IRAP when participants completed the  $C_{rel}$  IRAP first, but not when the order of completion was reversed (i.e., RCI IRAP first). Although this was an explorative study, the resulting data produced a tentative insight into the impact of response options employed in the IRAP and served to highlight the functional differences between  $C_{rel}$ s and RCIs as referenced in RFT literature.

It could be argued that this type of work fits with contemporary studies using the IRAP to investigate the dynamics of relational framing (e.g., Finn et al. 2016; Maloney and Barnes-

Holmes 2016; Finn et al. 2018;). In particular, Finn et al. (2016) looked at the potential impact of presenting rules prior to participant completion of consistent and inconsistent trial-types in the IRAP. Finn et al. (2018) examined in detail the effects of coherence of relations and stimuli including response options in IRAP trial-type presentations (e.g., DAARRE; differential arbitrarily applicable relational responding model). Such experimental work is necessary to feed into the wider arena (e.g., psychometric testing) that the IRAP is often used for (Finn et al. 2018). Consistent with behavioral research to date, this bottom-up approach when using the IRAP to investigate behavior, is not only essential but also fitting with the behavioral analytic perspective on which RFT is based. With this in mind, the current research built upon the work of Maloney and Barnes-Holmes (2016) to extend the IRAP technological research literature, and examined whether the type of response options used ( $C_{rel}$ s v. RCIs) affected on IRAP results. The  $C_{rel}$  IRAP used “Same” and “Opposite” as response options, and the RCIs IRAP used “Accurate” and “Inaccurate” as response options. The  $C_{rel}$ s were selected based on a review of the available empirical IRAP literature, which indicates that they have been frequently used in IRAP research and represent distinct bipolar  $C_{rel}$ s. The RCIs were selected as they are strict evaluative terms and are comparable to the RCI terms previously used in this line of IRAP research. The study was conducted with college students ( $N = 40$ ), and the order in which the type of IRAP was completed was counterbalanced across participants ( $n = 20$   $C_{rel}$  first;  $n = 20$  RCI first). Findings may help clarify whether these theoretical distinctions between the function of terms classed as  $C_{rel}$ s and RCIs is reflected in empirical experiments (Hayes and Barnes 1997). In particular, this study aimed to further examine if the terms used as response options in the IRAP affect the resulting data while exploring the dynamics of relational framing.

## Method

### Participants

A sample of 52 participants was recruited from the student population of the National University of Ireland (NUI), Maynooth for the current research. Each participant completed the experiment individually in quiet lab cubicles at the Department of Psychology of Maynooth University. Participants gave written informed consent to participate, with no penalty or obligation. No financial or other incentives were offered for participation in the study. Participants were asked to self-select if English was their primary language and their vision was normal or corrected to normal with prescribed lenses or glasses. They were advised to self-exclude if they had a history of seizures due to a negligible risk related to use

of a computer screen. The data from 12 participants who failed to maintain the predetermined IRAP performance criteria were removed prior to analysis. Thus, 40 participants ( $N = 40$ ; 17 males, 23 females) successfully completed the research procedures (age  $M = 23$  years, range 18–45 years).

### Apparatus/Materials

All participants completed both IRAPs on a laptop computer (Lenovo G50). The IRAP (2014) software, written in Microsoft Visual Basic 6.0, controlled the presentation of all stimuli and the recording of responses. Each trial of the IRAP presented one of two sample stimuli: “Pleasant” or “Unpleasant.” One of 12 target stimuli were also presented in each IRAP trial. The target stimuli consisted of six synonyms of the term “Pleasant” (good, positive, nice, likeable, lovely, and wonderful) and six synonyms of “Unpleasant” (bad, negative, nasty, unlikeable, horrible, and awful). Prior to commencing each block of trials, the IRAP program would present one of two “rules” that was to be followed by participants when responding on the subsequent trials. The rule presented before consistent trial-blocks was “Pleasant is Positive. Unpleasant is Negative.” The rule presented before inconsistent trial-blocks was “Pleasant is Negative. Unpleasant is Positive.”

### Procedure

At the top of the computer screen the IRAP presented one of two sample stimuli: either the word “Pleasant” or “Unpleasant” appeared along with a target word stimulus in the center of the screen such as “Likeable” or “Awful.” A set of response options (“Same” and “Opposite” on the  $C_{rel}$ s IRAP; “Accurate” and “Inaccurate” on the RCI IRAP) appeared in the bottom left- and right-hand corners of the screen (see Figure 1). Participants were informed that these positions would switch quasi-randomly from trial to trial (not appearing in the same position 3 times in succession). Participants were instructed to choose one of these response options for each trial, by pressing either the “d” or “k” key. All other keys were disabled during trials. The phrases “PRESS ‘d’ FOR \_\_\_\_\_” (e.g., “Same”) and “PRESS ‘k’ FOR \_\_\_\_\_” (e.g., “Opposite”) appeared directly above the onscreen response options on each trial, and these prompts indicated which key corresponded to which response option on said trial.

The instructions presented on screen by the IRAP program during this instructional phase of the study included the appropriate “rule,” which would indicate the feedback contingency for the upcoming block of trials. The “rule” alternated from block to block and participants were informed prior to each new block of trials that the previously correct and incorrect answers would now be reversed. The information was

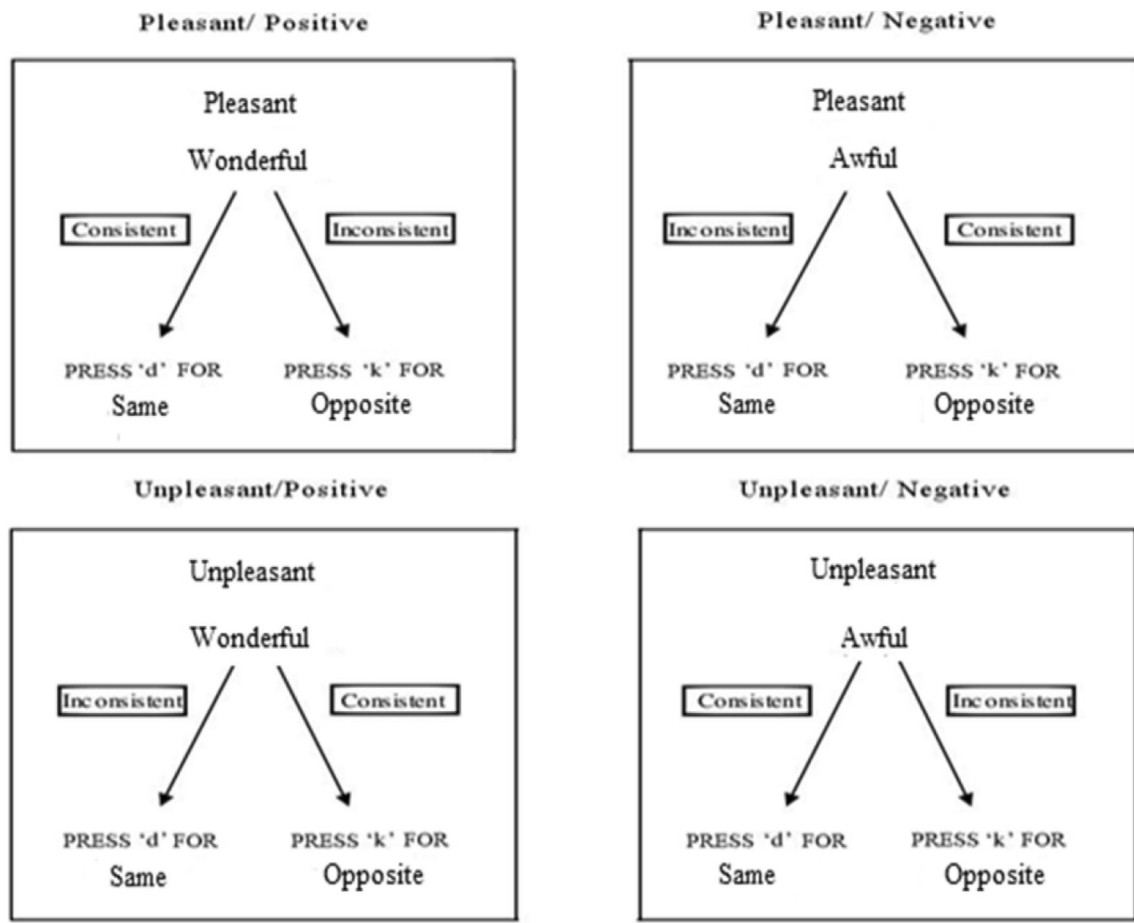


Fig. 1 Examples of the four IRAP trial types

provided that the upcoming trial-blocks were for practice, and the guidance to “Try to avoid the red ‘X’ on every question” was also presented. During subsequent test blocks, this guidance would change to “Please try to get as many right as possible.” In explaining the rule to participants they were informed that for some trial-blocks they would be required to respond in a manner consistent with the English language (e.g., when the rule stated that Pleasant is Positive and Unpleasant Negative, affirm those relations by selecting the appropriate response options) and that for some trial-blocks the task was to respond in a manner inconsistent with the English language (e.g., when the rule stated that Pleasant is Negative and Unpleasant is Positive, affirm those relations by selecting the appropriate response options). It was emphasized to participants that it was important for them to respond accurately, in accordance with the “rule,” and quickly, on all trial-blocks. The reasoning for this was not explained to participants until after the experimental process. Participants were informed that responding correctly (i.e., as designated) in a given trial would allow them to progress to the next trial and would remove all stimuli from the screen for a 400 ms interval until that new trial was presented. Incorrect responses would result in an emboldened red “X” appearing onscreen, which

would remain until the designated correct response had been emitted by the participant. If a participant failed to respond within 2,000 ms, an emboldened red exclamation mark would appear onscreen. The exclamation mark would remain there until the participant responded.

The IRAP procedure comprised of a minimum of eight blocks of trials: a minimum of two practice blocks followed by a fixed set of six test blocks. The same 24 trials were presented in each practice and test block. Within each block the sample stimuli “Pleasant” and “Unpleasant” were presented quasi-randomly across trials with each term appearing 12 times within the 24-trial block. The 12 target stimuli (e.g., Wonderful) were also presented in a quasi-random sequence with the constraint that each be presented twice across the 24 trials. Thus, each of the four possible trial-types (see Figure 1), were presented six times, though not twice in succession.

To progress from practice to test blocks participants were required to achieve an accuracy in responding of at least 80% with an average response latency of less than or equal to 2,000 ms. These criteria were used to produce an IRAP effect that may not be evident with less stringent criteria. If a participant failed to meet these criteria for either of the first two practice blocks, the level of responding they had achieved and the

standard required were presented on screen at the end of the set. Participants were permitted up to six further practice blocks to reach the required standard of responding. Participants who failed to meet these criteria after the additional practice were thanked for their participation, debriefed, and excused from the study. Their data were discarded. Successful participants proceeded to the six test blocks. During practice blocks the experimenter sat adjacent to the participant and watched as they responded to the IRAP.

The experimenter did not remain with the participant during test blocks and returned only after the task had been completed. Feedback was presented on screen immediately after each block of trials. This feedback detailed the accuracy and the median latency achieved for the trial-block. When ready to continue to the next block of trials, participants were told to hit the space bar. Upon completion of all six test blocks, participants were notified to alert the researcher via an onscreen message.

Each participant in this study was required to complete two consecutive IRAPs in one sitting. The sole difference between the two IRAPs was the type of response options used. The order in which participants completed both IRAPs was dependent upon their random allocation of to  $C_{rel}$  IRAP first ( $n = 20$ ) or RCI IRAP first ( $n = 20$ ). In each IRAP, 10 participants completed consistent trial-blocks first, whereas the other 10 completed inconsistent trial-blocks first.

## Results

### Data preparation

The data from a total of 12 participants were removed prior to data analysis. Four participants were unable to reach the criteria during initial practice blocks; an additional eight participants failed to maintain criteria on at least one of the two IRAPs.

The primary datum recorded by the IRAP was response latency. This is defined as the time in milliseconds that elapsed between the onset of a trial and the input of a correct response by a participant. In line with previous analyses of IRAP data, the response latency data for each participant were transformed into  $D$ -IRAP scores to control for the potential individual variations of responding which may confound when analyzing between group differences. This process of calculating  $D$ -IRAP scores is done by the IRAP program through the following steps: (1) response-latency data from all practice blocks are excluded from analysis; (2) trials with latencies exceeding 10,000 ms are removed from the dataset; (3) if in excess of 10% of test-block trial latencies were below 300 ms for a single participant dataset, the data of that participant was removed from analyses; (4) four standard deviations were calculated for the response latencies of test-blocks 1 and 2,

four more were calculated from test-blocks 3 and 4, and a further four were calculated from test-blocks 5 and 6, resulting in 12 standard deviations for the four trial types; (5) an average response-latency was calculated for the four trial types in each test block, resulting in 24 mean latencies; (6) by subtracting the mean latency of the pleasant-positive test-block from the mean latency of the corresponding pleasant-negative test block difference scores for each of the four trial types were calculated for each pair of test-blocks; (7) these difference scores were then divided by their corresponding standard deviation, calculated in step 4, producing one  $D$ -IRAP score for each trial-type for each pair of test blocks; (8)  $D$ -IRAP scores were calculated for each of the four trial types by averaging the scores for each trial-type across the three pairs of test blocks; (9) by averaging these 12 trial-type  $D$ -IRAP scores an overall  $D$ -IRAP score was calculated for each of the four trial-types. This method is an adaptation of the Greenwald, Nosek, and Banaji (2003)  $D$ -algorithm.

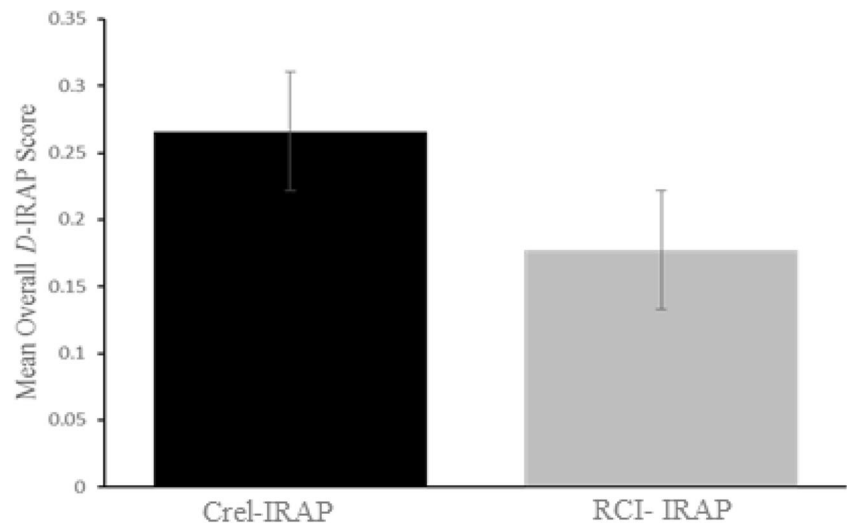
For participants who exceeded a response latency of 2,000 ms or who fell below 75% in accuracy on just one test block, analyses were conducted on the remaining two pairs of test blocks (as per Nicholson and Barnes-Holmes 2012). In particular, the test block pair that did not reach the required criteria was removed from analysis and the  $D$ -IRAP scores for that participant was recalculated. In the current study, 10 sets of participant data underwent this treatment prior to analysis. If a participant failed to reach criteria across two or more test block pairs, her or his entire dataset was removed from analysis.

### Data Analysis

Following data transformation, the overall  $D$ -IRAP scores for both IRAPs indicate that participants more rapidly affirmed consistent relations compared to inconsistent trial-blocks (see Figure 2 for graphic representation). In particular, the overall mean  $D$ -IRAP score for the  $C_{rel}$  IRAP was .27 ( $SD = .28$ ) and for the RCI IRAP the  $D$ -score was .18 ( $SD = .24$ ). Thus, the expected IRAP effect (bias favoring preexperimentally learned relations) was shown on both.

**Main analysis** Statistical analysis was conducted using a mixed between–within  $2 \times 2 \times 2 \times 4$  analysis of variance (ANOVA) on participant  $D$ -IRAP scores ( $N = 40$ ) to determine if type of response options, order of IRAP completion, or block-order (i.e., presenting consistent vs. inconsistent trial-blocks first) affected results across the four repeated IRAP trial-types. Both IRAPs showed the expected IRAP effect and participants more rapidly affirmed preexperimentally learned relations (Pleasant Positive, Unpleasant Negative). There was a main effect for type of response option used (Wilks's Lambda = .87,  $F(1, 36) = 5.59$ ,  $p = .02$ ,  $\eta_p^2 = .14$ ) and the IRAP effect was stronger in the  $C_{rel}$  IRAP (Figure 2).

**Fig. 2** Bar graph showing overall mean *D*-IRAP scores (with standard error bars) for both IRAPs. The graph indicates a pattern of responding biased towards congruent relations (i.e., Pleasant-Positive-Same, Pleasant-Positive-Accurate)



A statistically significant main effect was also shown for IRAP trial-type: Wilks's Lambda = .38,  $F(3, 34) = 18.37$ ,  $p < .001$ ,  $\eta_p^2 = .618$ . The mean overall *D*-IRAP scores for each of the four IRAP trial-type conditions across both IRAP types (C<sub>rel</sub> and RCI response options) are presented in Figure 3. There was no statistically significant two-way interaction shown between order of completion (e.g., C<sub>rel</sub> IRAP first or RCI IRAP first), and the type of response options used: Wilks's Lambda = .97,  $F(1, 36) = .976$ ,  $p = .33$ . There was, however, a statistically significant three-way interaction evident between the order of completion, the type of response option used, and the block-order presentation of consistent and inconsistent trial-blocks: Wilks's Lambda = .86,  $F(1, 36) = 5.66$ ,  $p = .02$ ,  $\eta_p^2 = .14$ .

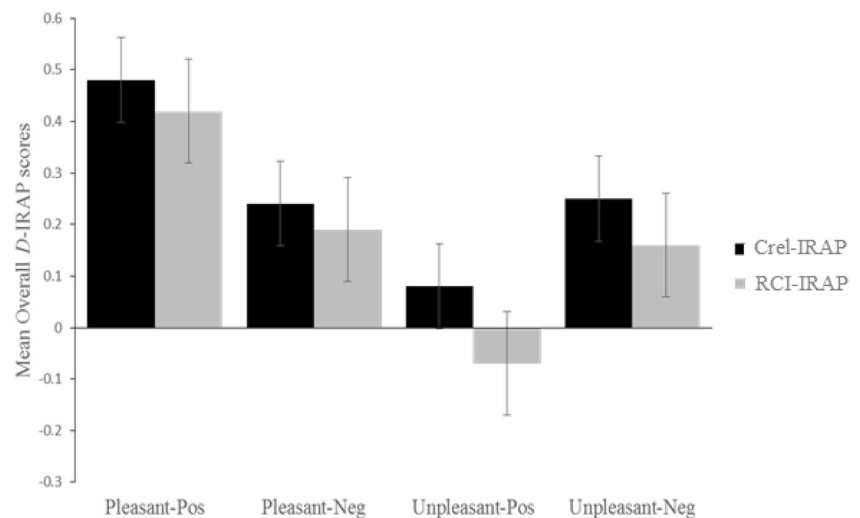
Eight one-sample *t*-tests were conducted to determine which of the four IRAP trial-types in both IRAPs were statistically significant. For both the IRAP types, the *D*-scores were statistically significant for all but the Unpleasant-Positive trial-type. Trial-type results for the C<sub>rel</sub> IRAP were as follows:

Pleasant-Positive ( $M = .48$ ,  $SD = .37$ ,  $F(39) = 8.21$ ,  $p < .001$ ); Pleasant-Negative ( $M = .24$ ,  $SD = .24$ ,  $F(39) = 3.70$ ,  $p < .001$ ); Unpleasant-Positive ( $M = .08$ ,  $SD = .46$ ,  $F(39) = 1.1$ ,  $p = .26$ ); Unpleasant-Negative ( $M = .25$ ,  $SD = .37$ ,  $F(39) = 4.36$ ,  $p < .001$ ). Trial-type results for RCI IRAP followed a similar pattern: Pleasant-Positive ( $M = .42$ ,  $SD = .38$ ,  $F(39) = 7.12$ ,  $p < .001$ ); Pleasant-Negative ( $M = .19$ ,  $SD = .39$ ,  $F(39) = 3.02$ ,  $p < .001$ ); Unpleasant-Positive ( $M = -.07$ ,  $SD = .41$ ,  $F(39) = -1.11$ ,  $p = .27$ ); Unpleasant-Negative ( $M = .16$ ,  $SD = .35$ ,  $F(39) = 2.95$ ,  $p < .001$ ).

## Discussion

Forty college students completed two similar IRAP procedures with differing response options (C<sub>rel</sub>'s "Same"/"Opposite" and RCIs "Accurate"/"Inaccurate"). Both IRAPs showed an IRAP effect as expected (i.e., participant responding was faster, on average, when affirming relations

**Fig. 3** The mean participant *D*-IRAP scores ( $N=40$ ), with standard error bars, for the four trial-types of both the C<sub>rel</sub> and RCI IRAP



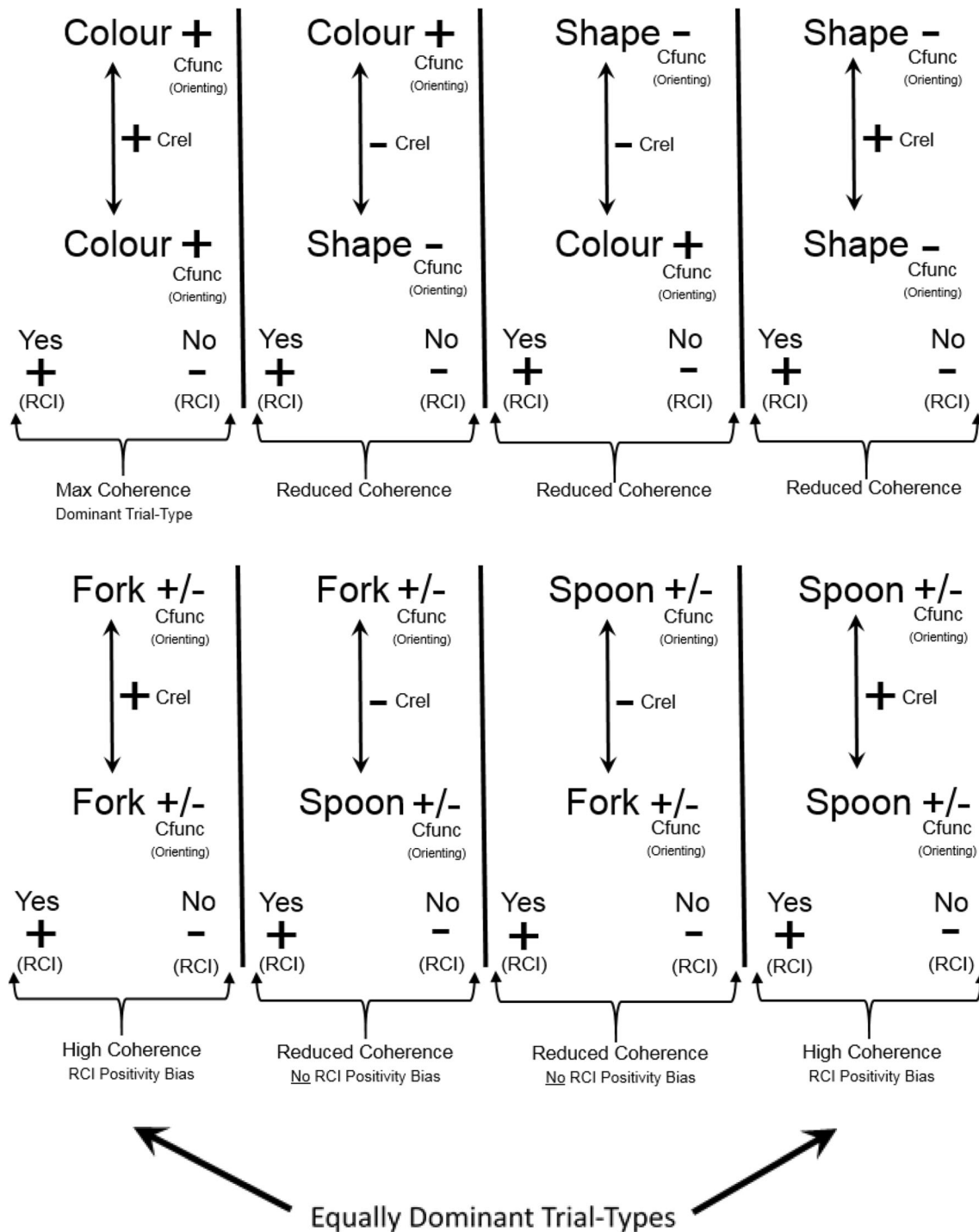
consistent with preexperimental learning, rather than converse "inconsistent" relations). A critical finding was that data analysis revealed that there was a statistically significant difference between responding across the two IRAPs. That is, the IRAP effect (bias) was shown to be stronger for the  $C_{rel}$  IRAP. Unlike in Maloney and Barnes-Holmes (2016) there was no interaction between IRAP type and order of completion ( $C_{rel}$  IRAP first vs. RCI IRAP first). There was a three-way interaction shown, however, among the type of response options, order of completion of IRAPs ( $C_{rel}$  or RCI first), and order of trial-block presentation (consistent vs. inconsistent first). Further analysis showed that the interaction applied to three of the four IRAP trial-types on both IRAPs, however, further research will be necessary for replication, clarification and interpretation purposes. This could be a spurious interaction or one that is connected to some artefact of the procedure.

Overall, the current results were consistent with the theory that  $C_{rel}$ s and RCIs may function differently (Hayes and Barnes 1997), and that the type of IRAP response options used in IRAP research may have an impact on IRAP effects shown. The larger effect size for the  $C_{rel}$  IRAP is in keeping with the DAARRE model (Finn et al. 2018). The model explains the Single-Trial-Dominance-Effect in IRAP trial-type 1 results in terms of greater coherence among Crels, Cfuncs, and stimuli. Trial-type 1 showed a greater difference compared to the other three trial-types, due to the greater coherence in the former and reduced coherence in the latter. DAARRE identifies three key sources of behavioral influence: (1) the relationship between the label and target stimuli (Crels); (2) the orienting functions of the label and target stimuli (Cfuncs); and (3) the coherence functions of the two response options (i.e., "Yes" or "No"). An example of stronger coherence in Crels provided by Finn et al. (see Figure 4) was Color–Color relations versus Color–Shape Relations. In this case, orienting functions in Color–Color relations were considered potentially stronger compared to Shape–Shape relations, due to greater frequency of the words "Color" versus "Shape" occurring in natural language. Trial-type 1 in the example had maximum coherence because in addition to Color–Color relations it involved selecting the positive response option ("Yes"). This was used to explain the Single-Trial-Dominance-Effect occurring in IRAP research where trial-type 1 showed a greater difference compared to the other three trial-types, which had reduced coherence. Finn et al. showed that when stimuli were manipulated to maximize coherence similarly in trial-types 1 and 4 (i.e., spoon–spoon relations and fork–fork relations), results showed similarly greater IRAP effects (two-trial-dominance) for both these trial-types compared to trial-types 2 and 3, which have reduced coherence.

Applying the DAARRE model in the current study, trial-type 1 would seem to have strong coherence between Crels in both IRAPs. Cfuncs and response options in trial-type 1 in both IRAPs are also coherent (i.e., Pleasant-Positive-Same/ Pleasant-Positive-Accurate). Trial-types 2 and 3 (Pleasant-Negative and Unpleasant-Positive, respectively) have reduced Crel coherence, as these relations would have been experienced less frequently if ever, in participants' preexperimental verbal histories. Trial-type 4 (Unpleasant-Negative relations) would have coherence in Crels, but possibly reduced orienting functions due to negativity as per the Finn et al. (2018) example. Thus, trial-type 1 in both IRAPs could be said to have maximum coherence compared with the other three trial-types, and accordingly showed stronger IRAP effects (single-trial-type-dominance effects), as greater coherence facilitates greater speed of responding in participants (i.e., shorter response latencies). The effect was shown across both IRAPs, but there was a statistically significant difference between results of IRAPs, and the IRAP effect was stronger in the Crels IRAP versus the RCI IRAP. It is possible that Pleasant-Positive-Same relations (Crel IRAP) are more in line with participants' experience of natural language, compared to Pleasant-Positive-Accurate (RCI IRAP), thus, although speculative, a potential interpretation is that trial-type 1 in the Crel IRAP may have had greater coherence, which influenced outcomes.

Findings of greater IRAP effects with Crels versus RCIs in the current research are consistent with previous findings (Maloney and Barnes-Holmes 2016), however, these findings are preliminary indications that different types of IRAP response options should not be assumed to be functionally equivalent; more research is warranted, however, before data could be considered conclusive. It should be noted also that Maloney and Barnes-Holmes (2016) did not identify a main effect for response options as in the current study, however this study was not a direct replication. For example, the response options used currently in the Crel IRAP were more polarized (Same/Opposite) compared to those used in Maloney and Barnes-Holmes (2016; i.e., Same/Different) which again could have affected coherence and in turn may influence results.

A potential concern was that after completion a small number of participants reported in discussions with the experimenter that the similarity of the RCIs used ("Accurate" vs. "Inaccurate") made the distinction between them a little difficult when trying to respond rapidly (although they expressed a preference for the RCIs compared to the  $C_{rel}$  response options). Future empirical analysis could be implemented to confirm or disconfirm if RCIs that are topographically similar except for a negative prefix such as "in" or "un," would affect participant responding on the IRAP. For example, whether response



**Fig. 4** The DAARE model as it applies to the shapes and colors stimulus set (upper panel) and the forks and spoons stimulus set (lower panel). Courtesy of Finn et al. (2018)

options “True/Untrue” would exert differential effects on participant responding compared to “True/False” response options. If response options such as True/Untrue, Accurate/Inaccurate, are less discriminable (reported anecdotally in the current study), it seems plausible that this could result

in longer response latencies (slower responding) across both trial-blocks, which could obscure any IRAP effect that might be shown in conditions more conducive to rapid responding (more readily coherent in DAARE terms). In REC terms, taking more time to respond could potentially



diminish the BIRR quality of participant responding (i.e., automatic responding), which is found when time pressure is applied, and conversely, could facilitate some deliberation, or EERR-type responding.

To summarize, the results were consistent with the theory that the type of IRAP response options used in the procedure ( $C_{rel}$  vs. RCI) may have an impact on IRAP effects demonstrated. Moreover, the three-way interaction shown (type/order of trial-type) could be viewed as an indicator that multiple aspects of the IRAP methodology are potentially relevant to participant performance. Current findings appear to fit well with the DAARRE model of coherence and affect on single-trial-type dominance effects. Concerning the issue of conclusive results and recommendations to IRAP researchers, first, a direct replication of the current study and/or Maloney and Barnes-Holmes (2016) may be necessary for extended systematic analysis and replication of effects. Otherwise, consideration of a rationale for using  $C_{rels}$  or  $C_{funcs}$  as response options in IRAP research may be warranted, and perhaps particular caution should attach to the use of IRAP response options that are topographically similar except for a negative prefix, until potential impacts are further elucidated.

**Availability of Data and Materials** All data and materials used in this study are available on request from the corresponding author.

**Funding Information** The second author is funded by the Marie-Sklodowska-Curie Actions COFUND Collaborative Research Fellowship for a Responsive and Innovative Europe (CAROLINE).

## Compliance with Ethical Standards

**Conflict of Interest** On behalf of all authors, the corresponding author states that there is no conflict of interest.

**Ethical approval** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

**Informed consent** Informed consent was obtained from all individual participants included in the study.

## References

- Barnes-Holmes, D., Hayes, S. C., Dymond, S., & O'Hara, D. (2001). Multiple stimulus relations and the transformation of stimulus functions. *Relational Frame Theory*, 51–71. [https://doi.org/10.1007/0-306-47638-X\\_3](https://doi.org/10.1007/0-306-47638-X_3)
- Barnes-Holmes, D., Barnes-Holmes, Y., Power, P., Hayden, E., Milne, R., & Stewart, I. (2006). Do you really know what you believe? Developing the Implicit Relational Assessment Procedure (IRAP) as a direct measure of implicit beliefs. *The Irish Psychologist*, 32, 169–177.
- Barnes-Holmes, D., Murphy, A., Barnes-Holmes, Y., & Stewart, I. (2010). The Implicit Relational Assessment Procedure (IRAP): Exploring the impact of private versus public contexts and the response latency criterion on pro-white and anti-black stereotyping among white Irish individuals. *The Psychological Record*, 60, 57–66.
- Barnes-Holmes, D., Waldron, D., Barnes-Holmes, Y., & Stewart, I. (2009). Testing the validity of the Implicit Relational Assessment Procedure and the Implicit Association Test: Measuring attitudes toward Dublin and country life in Ireland. *The Psychological Record*, 59, 389–406.
- Campbell, C., Barnes-Holmes, Y., Barnes-Holmes, D., & Stewart, I. (2011). Exploring screen presentations in the Implicit Relational Assessment Procedure (IRAP). *International Journal of Psychology & Psychological Therapy*, 11, 377–388.
- Cullen, C., & Barnes-Holmes, D. (2008). Implicit pride and prejudice: A heterosexual phenomenon? In M. A. Morrison & T. G. Morrison (Eds.), *The psychology of modern prejudice* (pp. 195–223). New York: Nova Science Publishers.
- Dawson, D. L., Barnes-Holmes, D., Gresswell, D. M., Hart, A. J., & Gore, N. J. (2009). Assessing the Implicit Beliefs of Sexual Offenders using the Implicit Relational Assessment Procedure: A first study. *Sexual Abuse*, 21, 57–75.
- Finn, M., Barnes-Holmes, D., Hussey, I., & Graddy, J. (2016). Exploring the behavioural dynamics of the implicit relational assessment procedure: The impact of three types of introductory rules. *The Psychological Record*, 66, 309–321. <https://doi.org/10.1007/s40732-016-0173-4>.
- Finn, M., Barnes-Holmes, D., & McEnteggart, C. (2018). Exploring the single-trial-type-dominance-effect in the IRAP: Developing a Differential Arbitrarily Applicable Relational Responding Effects (DAARRE) model. *The Psychological Record*, 68, 11–25.
- Golijani-Moghaddam, N., Hart, A., & Dawson, D. L. (2013). The Implicit Relational Assessment Procedure: Emerging reliability and validity data. *Journal of Contextual Behavioral Science*, 2, 105–119.
- 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(2), 197–216. <https://doi.org/10.1037/0022-3514.85.2>
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality & Social Psychology*, 74(6), 1464–1480.
- Hayes, S. C., & Barnes, D. (1997). Analysing derived stimulus relations requires more than a concept of stimulus class. *Journal of the Experimental Analysis of Behavior*, 68, 225–233.
- Hayes, S. C., Barnes-Holmes, D., & Roche, B. (2001). *Relational Frame Theory: A Post-Skinnerian account of human language and cognition*. New York: Plenum Press.
- Hussey, I., & Barnes-Holmes, D. (2012). The Implicit Relational Assessment Procedure as a measure of implicit depression and the role of psychological flexibility. *Cognitive & Behavioral Practice*, 19, 573–583.
- 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*, 65, 731–742.
- Maloney, E., & Barnes-Holmes, D. (2016). Exploring the behavioural dynamics of the Implicit Relational Assessment Procedure: The role of relational contextual cues versus relational coherence indicators as response options. *The Psychological Record*, 66, 395–403.
- Murphy, C., MacCarthaigh, S., & Barnes-Holmes, D. (2014). Implicit Relational Assessment Procedure and attractiveness bias: Directionality of bias and influence of gender of participants. *International Journal of Psychology & Psychological Therapy*, 14, 333–351.

- McEnteggart, C., Barnes-Holmes, Y., & Adekuoroye, F. (2016). The effects of a voice hearing simulation on implicit fear of voices. *Journal of Contextual Behavioral Science*, *5*, 154–159.
- McKenna, I. M., Barnes-Holmes, D., Barnes-Holmes, Y., & Stewart, I. (2007). Testing the fake-ability of the Implicit Relational Assessment Procedure (IRAP): The first study. *International Journal of Psychology & Psychological Therapy*, *7*, 123–138.
- Nicholson, E., & Barnes-Holmes, D. (2012). Developing an implicit measure of disgust propensity and disgust sensitivity: Examining the role of implicit disgust propensity and sensitivity in obsessive-compulsive tendencies. *Journal of Behaviour Therapy & Experimental Psychiatry*, *43*, 922–930.
- 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 & Emotion*, *27*, 1441–1449.
- Sidman, M. (1971). Reading and auditory-visual equivalences. *Journal of Speech & Hearing Research*, *14*, 5–13.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.