Exploring the Behavioral Dynamics of the Implicit Relational Assessment Procedure: The Impact of Three Types of Introductory Rules

Martin Finn, Dermot Barnes-Holmes, Ian Hussey

Department of Experimental, Clinical and Health Psychology, Ghent University, Belgium

and

Joseph Graddy

University of Waikato, New Zealand

Corresponding author:

Martin Finn

Department of Experimental, Clinical and Health Psychology

Henri Dunantlaan 2

9000 Ghent, Belgium

Email: Martin.Finn@UGent.be

Abstract

The Implicit Relational Assessment Procedure (IRAP) is increasingly used in applied and clinical settings, and yet many of the procedural variables of the measure have not been subjected to a systematic analysis. One such variable is the type of rules that are employed when instructing the IRAP and the effect this might have on participants' performances. In the current article, three experiments were used to assess the impact of three different types of rules or instructions on IRAP performance. The instructions varied in the degree to which they specified parts of the relational network being assessed by the IRAP. The findings showed that the type of rule presented to participants during an IRAP can have a dramatic effect on the strength and direction of the trial-type effects that are produced by the measure. Furthermore, the type of instructions employed appear to interact with the order in which the IRAP blocks are presented (history-consistent versus history-inconsistent). The findings indicate that the behavioral dynamics that occur when participants complete an IRAP require extensive and systematic experimental and conceptual analyses, and this work will likely have an important bearing on research seeking to investigate the predictive validity of the IRAP in applied research settings.

Relational Frame Theory (RFT) has been offered as a modern behavior-analytic account of human language and cognition (Hayes, Barnes-Holmes, & Roche, 2001). The basic idea is that human language is composed largely of different patterns of relational responses, defined as relational frames. Demonstrating such patterns of relational responding in the research laboratory typically involves training or reinforcing a minimal set of relational responses (e.g., A *same as* B and B *same as* C) and then testing for derived relational responses (e.g., A *same as* C; C *same as* A). If the predicted relational responses emerge the stimuli are said to participate in a relational frame (in this case, the frame of coordination). According to RFT, relational frames may combine into what are called relational networks, which are defined as coherent when they are generally consistent with previous histories of relational responding (see Barnes-Holmes, Hayes, Dymond, & O'Hora, 2001). A large number of studies have explored this theory of human language and cognition (see Dymond & Roche, 2013; Hughes & Barnes-Holmes, in press, for recent reviews), and the evidence thus far is largely supportive.

In more recent years, some researchers have shifted their attention from testing the basic RFT model to analyzing the relative strength of relational responding as defined by RFT (see Barnes-Holmes, Barnes-Holmes, Hussey, & Luciano, in press; Hussey, Barnes-Holmes, & Barnes-Holmes, 2015). The development of the Implicit Relational Assessment Procedure (IRAP), which grew directly out of the theory, was instrumental in pursing this line of research (see Hughes & Barnes-Holmes, 2011). The IRAP is a computer-based task which presents label stimuli at the top of the screen, such as "FLOWER" and "INSECT", and target stimuli which appear in the middle of the screen, such as "PLEASANT", "GOOD", "UNPLEASANT", and "BAD". Which labels and targets appear on screen at any point in the task is quasi-random, with the constraint that the two types of labels will appear with the two types of targets an equal number of times within a pre-determined block of trials. The labels

and targets thus form a 2x2 cross-over design that yields four different trial-types on the IRAP (in the current example: *Flower-Positive; Flower-Negative; Insect-Positive;* and *Insect-Negative*). On each trial, two response options are provided, which require participants to confirm or deny specific relationships between the label and target stimuli. For example, "FLOWER" and "PLEASANT" might appear on a given trial with the response options "True" and "False", and in this case participants would be required to confirm (i.e., pick "True") or deny (i.e., pick "False") that flowers are pleasant.

The IRAP operates by requiring opposing patterns of responding across successive blocks. For example, "FLOWER" and "PLEASANT" would require the response "True" on one block and "False" on the next block. The IRAP operates on the assumption that, all things being equal, history-consistent response patterns will be emitted more readily than history-inconsistent patterns (Barnes-Holmes, Barnes-Holmes, Stewart, & Boles, 2010). Consider for example that certain relational responses, within the verbal community, are more likely to be reinforced than punished (e.g. affirming that flowers are pleasant), while others are more likely to be punished than reinforced (e.g. denying that flowers are pleasant). Thus, the more readily emitted pattern of responding on an IRAP may be seen as indicative of the natural contingencies operating in the wider verbal community. Broadly speaking, the IRAP is scored by subtracting the mean response latency for one pattern of responding from the mean response latency of the opposite pattern. Any resultant difference is deemed to be reflective of the differential behavioral histories involved in the two patterns of responding (note that the difference score is typically normalized using an appropriate statistical method, such as a derivative of Cohen's *d*).

In considering the types of effects that have been obtained on the IRAP, behavioral researchers have often referred to brief and immediate relational responses (BIRRs), which are emitted relatively quickly within a short window of time after the onset of the stimuli

presented on any given IRAP trial. In contrast, extended and elaborated relational responses (EERRs) are more complex and are seen as being emitted more slowly and as such occur over a longer period of time. The distinction between BIRRs and EERRs was first formalized in the context of the Relational Elaboration and Coherence (REC) model, which was offered as an initial RFT approach to implicit cognition (Barnes-Holmes et al., 2010; Hughes, Barnes-Holmes, & Vahey, 2012). The basic idea behind the model is that the types of effects observed on the IRAP, and indeed other implicit measures, were due to the fact that the task targeted BIRRs rather than EERRs. For example, the fact that the IRAP requires participants to respond relatively quickly on each trial, almost by definition, encourages participants to emit BIRRs during exposure to the procedure. The relative strength or probability of such BIRRs is deemed to be related to functionally similar stimuli in the participants' behavioural histories.[1]

Imagine, for example, a white individual who has resided exclusively in white neighborhoods, has no non-white friends or family members, and has been exposed to many media images of black people as violent drug dealers and inner-city gang members. When presented with an IRAP with pictures of black males carrying guns it is likely, according to the REC model, that BIRRs for confirming that black men are "dangerous" and "criminals" may be more probable than denying such relations. As a result, the participant may respond more rapidly on the IRAP when required to confirm, rather than deny, that a black man carrying a gun is dangerous. In effect, an anti-black racial bias may be revealed by the IRAP. In contrast, such a bias might be absent if the same participant was asked to rate the pictures of the black men from the IRAP with no time constraints for doing so. The lack of racial bias in the latter context is explained in the REC model by appealing to EERRs, which occur given

[1] We would argue that BIRRs and EERRs are relativistic concepts and are best seen as lying on a continuum. At one end of the continuum responses are extremely brief and/or immediate and at the other end extremely extended and/or elaborated. As argued by Barnes-Holmes, et al. (2010, p. 537), according to the REC model the IRAP targets relational responses that are more BIRR-like than EERR-like.

sufficient time for an individual to respond in accordance with a relationally coherent network. In the context of the current example, the participant might fail to report any initial BIRR that involves perceiving the pictures of black males as "dangerous" based on additional relational responding, such as "It is wrong to discriminate on the basis of race" and "I am not a racist", etc. In general, therefore, the REC model attempts to explain the emergence of specific response biases on the IRAP by arguing that the procedure tends to reveal BIRRs rather than EERRs.

In concluding that the IRAP reveals BIRRs rather than EERRs, the REC model assumes that this applies, with roughly equal force, to all four trial-types. In other words, the IRAP is seen as providing a measure of the strength or probability of four functionally independent BIRRs. Imagine, for example, an IRAP that aimed to assess the response probabilities of four well-established verbal relations pertaining to non-valenced stimuli such as shapes and colors. Across trials, the two label stimuli, "Color" and "Shape" could be presented with target words consisting of specific colors ("Red", Green", and "Blue") and shapes ("Square", "Circle", and "Triangle"). As such, the IRAP would involve presenting four different trial-types that could be designated as (i) *Color-Color*, (ii) *Color-Shape*, (iii) *Shape-Color* and (iv) *Shape-Shape*. One may conceive of these four trial-types as participating in a single relational network as depicted in Figure 1. According to the REC model, responding in accordance with this network, such as describing each of the relations involved, may be seen as an EERR because it would require a relatively elaborated and extended relational response pattern. As such, each element within that pattern of relational responding would be functionally dependent on the other elements. For example, saying that "a square is a shape and red is a color" would cohere with saying that "red is not a shape and a square is not a color." In other words, responding in accordance with such relational networks requires, by definition, that coherence within the network is maintained. Maintaining such

coherence, particularly as networks become more extended and elaborated, requires time. In contrast, an IRAP is completed under time pressure and thus the functional dependencies or coherence within a given network may not be maintained.

To appreciate this point, consider again the nature of the IRAP itself. During a Shapes-and-Colors IRAP, participants would be required to respond in a manner that was consistent with their pre-experimental histories during some blocks of trials; (i) *Color-Color-True*; (ii) *Color-Shape-False*; (iii) *Shape-Color-False*; and (iv) *Shape-Shape-True*. On other blocks of trials, the participants would have to respond in a manner that was inconsistent with those histories; (i) *Color-Color-False*; (ii) *Color-Shape-True*; (iii) *Shape-Color-True*; and (iv) *Shape-Shape-False*. All things being equal, the REC model assumes that the IRAP requires participants to emit BIRRs (rather than EERRs) on each trial. Thus, when the four trial-type effects are calculated, by comparing response latencies for history-consistent with history-inconsistent blocks of trials, functionally independent effects would be observed for each trial-type. For example, when a participant responds, *Shape-Shape*-True on a given trial, he or she is not required to cohere that response with other elements within the network on that particular trial (because there is, in principle, insufficient time to do so). Or more informally, when an individual responds on any given trial in an IRAP the REC model assumes that it is not possible to "work through" or "rehearse" every other element within that network to ensure overall relational coherence. The current study constitutes a first step in testing this basic assumption by providing participants with different types of rules on how to respond during exposure to a Shapes-and-Colors IRAP.

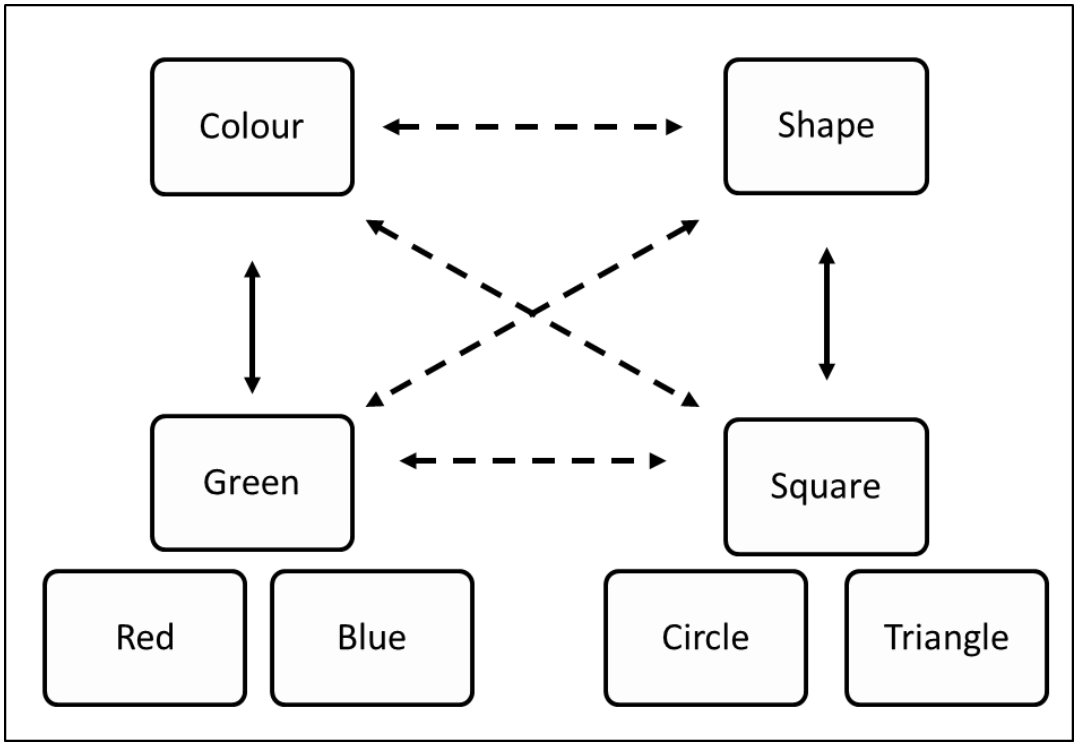

Figure 1. Representation of the relational network that is assessed by a shapes-and-colors IRAP. Dashed lines indicate a relation of difference between stimuli, whereas solid lines indicate relations of coordination between stimuli.

To appreciate the rationale for the research reported herein, consider again the basic REC model assumption that on any given IRAP trial it is not possible, due to time constraints, to check for relational coherence with the network that constitutes all four trial-types. In practice, however, this may not be the case. If a participant is capable of completing a private relational response on an IRAP trial within approximately 800-1000ms, there is an additional 1000-1200ms available for relational activity that extends beyond the targeted response (see Hughes, Hussey, Corrigan, Jolly, Murphy, & Barnes-Holmes, 2016). Crucially, if participants were encouraged to engage in such additional relational activity, perhaps through the provision of specific rules, this may well impact upon the IRAP performance. For example, imagine that participants were instructed on a Shapes-and-Colors IRAP to "respond as if colors are colors and shapes are shapes". If participants emitted this rule (privately) on many if not most of the trials before emitting a response it may reduce the BIRR-like properties of

such responses, and thus the REC model would no longer predict relatively strong IRAP effects for all four trial-types (see Barnes-Holmes, et al., 2010, p. 537). The three studies reported here aimed to test this suggestion.

Experiment 1

The first study aimed to test the extent to which specific versus general rules presented before each block of trials in an IRAP impacted upon the relative sizes of the four trial-type effects. Specifically, in Experiment 1 participants were provided with two types of rules before completing each block of a Shapes-and-Colors IRAP. Half of the participants were presented with a "specific" rule that specified part of the "shapes and colors" relational network, targeting in particular two of the trial-types, and half of the participants were presented with a "general" rule that did not specify specific parts of the network or specific trial-types. This initial experiment was largely exploratory in nature although some differences in the pattern of IRAP effects across the two types of rules (specific vs. general) were expected.

Before continuing, it seems important to note that on the grounds of intellectual honesty we feel that it would be inappropriate to provide a *detailed* theoretical analysis of the rationale for exploring the impact of detailed versus general rules on IRAP performances. On balance, the decision to compare specific versus general rules was not a complete "stab in the dark." Rather, we suspected that providing rules that specified particular parts of the relational network targeted by the IRAP may encourage the type of additional relational activity noted above, and thus reduce the BIRR-like properties of some of the relational responding. In contrast, providing general rules that did not specify particular parts of the network may be less likely to encourage the additional relational activity and thus relatively consistent BIRR-like responding would be observed. As will become clear, the results of the current experiment (and two follow-up experiments) reported here provide some evidence to support

the foregoing general conceptual analysis. A more detailed theoretical analysis was made possible in light of the data gathered across all three experiments, and this is presented in the General Discussion.

Method

Participants

Sixty-two undergraduate students of psychology at Maynooth University participated in the experiment as part of a required practical element of their degree course. All participants completed an IRAP and a valence questionnaire. The sample comprised of 42 females and 20 males with an age range 17-43 yrs. The participants were randomly assigned to one of two groups, labelled here as the general-rule (GR) group (N = 32), and the specific-rule (SR) group (N = 30).

Materials

IRAP. The IRAP was presented on standard personal computers. The IRAP software was used to present the instructions, the stimuli, and to record responses. Each trial presented one of two labels; "Color" or "Shape." The label stimulus was presented along with one of six target stimuli. The target stimuli were all words, three denoting colors, "Red," "Green," and "Blue" and the other three denoting shapes, "Triangle," "Circle," and "Square." The rules provided to the GR-group were; (1) "Respond correctly to the stimuli," and (2) "Respond incorrectly to the stimuli." The rules provided to the SR-group were; (1) "Respond as if colours are colours and shapes are shapes," and (2) "Respond as if shapes are colours and colours are shapes."

Questionnaire and rating scale. A questionnaire comprising of demographic information (age, gender) and a rating scale for the stimuli presented in the IRAP were completed by all participants. The rating instrument comprised of a seven-point scale ranging from "very negative" to "very positive." Each of the labels and targets used in the IRAP were

rated. Data from the rating scales were analyzed in this and subsequent experiments, but as expected nothing of any importance emerged (e.g., no significant differences between the two rule conditions) and thus these analyses are not reported here.

Procedure

Details of the group-based session. The experiment was completed in two group-based sessions in the computer laboratory in the Department of Psychology at Maynooth University. The laboratory contained 33 desktop computers with an IRAP program installed on each machine. The experiment was run across two one-hour sessions, with the first session scheduled between 2-3pm and the second session scheduled for 3-4pm. Participants attending the 2-3 session received the general rules and participants attending the 3-4 session received the specific rules. Upon entering the computer laboratory, participants were asked to select a computer to sit at and to wait for further instruction. When all participants for that session were seated at a computer, the researcher commenced the experiment.

The researcher welcomed the participants and informed them that he would present a block of the IRAP program via the overhead projector so that everyone would have the opportunity to see what they would be asked to do during the experiment. He then proceeded to present a block of trials from the IRAP while describing what responses were required and the feedback contingencies that would apply. Thus, for example, a trial might have been presented in which the word “Shape” appeared as a label and the word “Circle” appeared as a target. The researcher pointed to these two stimuli and then stated that in this case responding “True” rather than “False” would be required, and to do this pressing the “D” key (rather than “K”) was appropriate. The researcher then noted that when this response was emitted the stimuli disappeared from the screen and were replaced almost immediately with the stimuli for another trial. The researcher continued to present a block of trials in this way providing examples of responding correctly and incorrectly within that block, noting for example, that

an incorrect response produced a red X in the middle of the screen and the program only continued to the next trial when the correct response was emitted.

When the researcher had worked through a single block of trials on the overhead projector, he invited the participants to click the mouse for their particular computer and follow the instructions that were presented on screen. Participants were also instructed to turn off mobile phones and any other electronic devices that might distract them from the experiment and to refrain from talking until everyone had completed the experimental session. A number of volunteer graduate students remained in the room throughout the session to ensure that all participants complied with this instruction.

The IRAP. On each trial of the IRAP, four words appeared on screen; a label at the top center of the screen (“Color”, or “Shape”); a target at the center of the screen (“Red”, “Green”, “Blue”, “Circle”, “Square”, or “Triangle”), and the two response options “True” and “False” at the bottom left and right of the screen, respectively. 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 (the words, “True” and “False”) 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. Examples of each type of trial to which participants were exposed are shown in Figure 2.

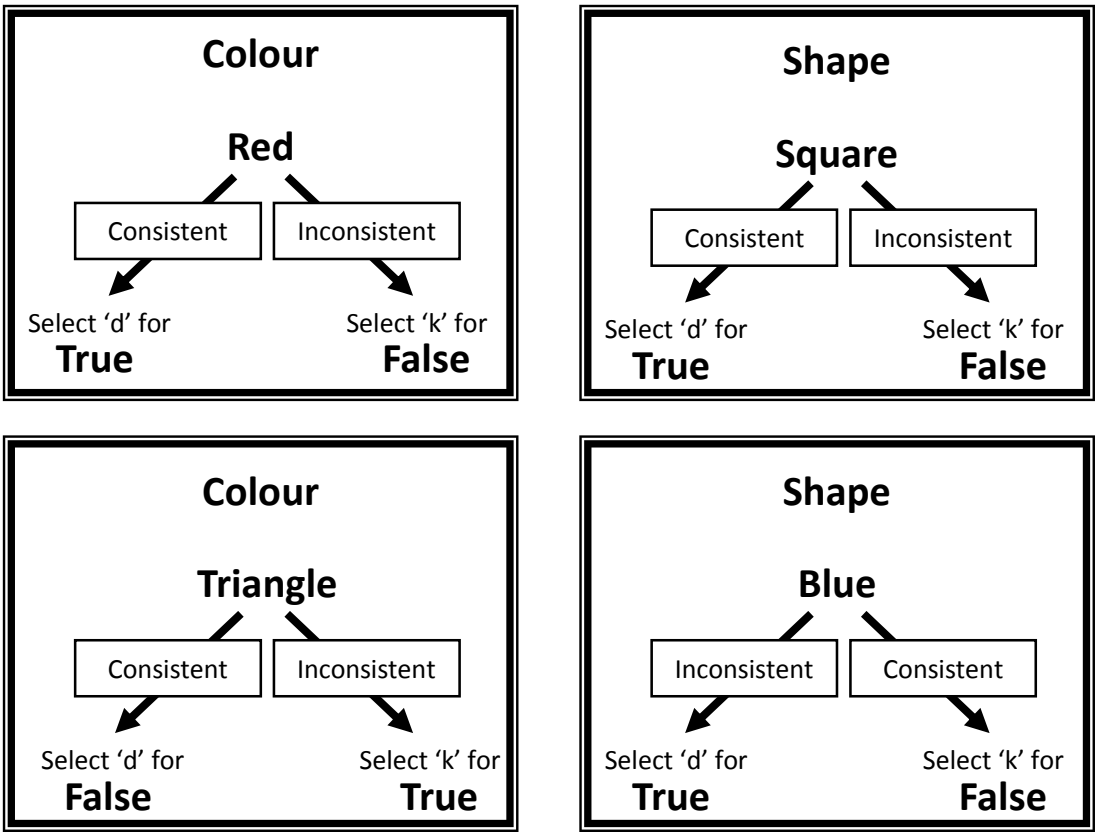

Figure 2. Diagrammatic representation of the four IRAP trial-types presented to participants. Arrows and boxes containing the words "consistent" and "inconsistent" did not appear on screen.

When participants selected the response option that was deemed correct within that block of trials the label, target, and response option stimuli were removed immediately from the screen for an inter-trial interval of 400 ms, after which the next trial was presented (i.e., a label, target, and two response options appeared simultaneously). When participants selected the response option that was deemed incorrect for that block of trials the stimuli remained on screen and a red X appeared beneath the target stimulus. The participants were required to select the correct response option, and only then did the program proceed directly to the 400 ms inter-trial interval (followed immediately by the next trial). This pattern of trial presentations, with corrective feedback, continued until the entire block of 24 trials was presented. The IRAP program presented the trials in a quasi-random order within each block with the constraint that each label was presented twice with each target stimulus across the 24

trials. Consistent with the majority of previously published IRAP studies, the trials presented within each block may be described as consisting of four different trial-types. In the current study, the four different combinations of label and target stimuli may be denoted as (i) *Color-Color*, (ii) *Color-Shape*, (iii) *Shape-Color*, and (iv) *Shape-Shape* (see Figure 2).

When participants completed a block of trials, the IRAP program provided them with feedback on their performance during that block. The feedback consisted of a message informing them how accurately they had responded in terms of percentage correct and how quickly they had responded in terms of median response latency. The latter was calculated from stimulus onset to the first correct response across all 24 trials within the block. Participants were required to achieve a minimum accuracy of 80 percent correct and a maximum median latency of no more than 2000 ms on each block within a pair. The IRAP program was set to allow participants up to 4 pairs of practice blocks to achieve these criteria. Only when participants achieved these criteria across both Blocks 1 and 2, or Blocks 3 and 4, or Blocks 5 and 6, or Blocks 7 and 8 were they permitted by the IRAP program to continue to the critical test blocks.

A fixed set of six test blocks were presented with no accuracy or latency criteria in order 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 (see below). In addition to the accuracy and latency feedback the IRAP program provided participants with tailored feedback instructions depending on their level of performance in the previous block. Participants who failed to achieve both accuracy and latency criteria were presented with the following text on screen between blocks “Learn to accurately follow the rule before attempting to respond quickly.” If the accuracy criterion had been achieved but not the latency criterion, the program presented the instruction “Continue responding as accurately as you

can. You'll naturally go quickly when your responses are accurate." If both latency and accuracy criteria had been met, the program presented the following instruction "Continue responding both as accurately and quickly as you can." These instructions were presented following both practice and test blocks.

Two types of feedback contingencies were applied across the practice and test blocks of the IRAP, denoted here as consistent versus inconsistent with the natural contingencies operating in the wider verbal community. The contingencies deemed consistent required participants to choose "True" on *Color-Color* and *Shape-Shape* trial-types and to choose "False" on *Color-Shape* and *Shape-Color* trial-types. The contingencies deemed inconsistent required participants to respond in an opposite pattern, choosing "False" on *Color-Color* and *Shape-Shape* trial-types and choosing "True" on *Color-Shape* and *Shape-Color* trial-types. The IRAP program typically applies the feedback contingencies in one of two patterns. For one pattern, the first block and all subsequent odd numbered blocks employed the consistent feedback; the second and all subsequent even numbered blocks employed the inconsistent feedback. For the second pattern, the first block and all odd numbered blocks employed the inconsistent feedback, and the second and all even numbered blocks employed the consistent feedback. The use of these two patterns of feedback contingencies was counterbalanced (approximately) across the participants in the current study (i.e., by pre-setting the IRAP program across the computers accordingly). In other words, half of the participants in each of the two rule conditions were presented with an IRAP that commenced with consistent feedback and then alternated from inconsistent to consistent across all subsequent blocks; the other half were presented with an IRAP that commenced with inconsistent feedback and then alternated from consistent to inconsistent across blocks thereafter.

As noted previously, participants were divided into two groups, the GR and SR groups. For both groups two rules were presented, and these were determined by the feedback

contingencies that were employed in the immediately following block of trials. For the GR group, each block of trials that employed the consistent feedback presented the rule: "Respond correctly to the stimuli" and each block that employed the inconsistent feedback presented the rule "Respond incorrectly to the stimuli". In other words, the rules informed participants how to respond during the next block of trials in a way that would avoid the red X, without specifying exactly how to do so. For the SR group, each block of trials that employed the consistent feedback presented the rule: "Respond as if colors are colors and shapes are shapes" and each block that employed the inconsistent feedback presented the rule "Respond as if shapes are colors and colors are shapes". In this case, therefore, the rules informed participants how to respond during the next block of trials by specifying exactly what pattern of responding to each trial-type was required to avoid the red X. Upon completion of the IRAP, all participants proceeded to the questionnaire. Thereafter participants were thanked for their time, debriefed, and dismissed.

Results and Discussion

The primary datum of the IRAP is the response latency defined as the length of time in milliseconds from stimulus presentation to a correct response on a particular trial. If participants maintained the accuracy and latency performance criteria across all six test blocks the data from all blocks were included in the analyses. If, however, a participant failed to maintain the criteria on one or both blocks within a given test-block pair (blocks 1 & 2; blocks 3 & 4; blocks 5 & 6), the data from that pair of test blocks were removed from the analyses. If a participant failed to maintain the criteria on two or more blocks from different test block pairs all of the data from that participant were removed. This practice was similar to that adopted by Nicholson and Barnes-Holmes (2012) in order to avoid excessively high attrition rates. Application of these criteria resulted in the exclusion of 20 of the 62 participants from the analysis, which is relatively high but may be accounted for, at least in part, by the use

relatively brief and rigidly set instructions. Note also, that attrition did not differ dramatically across the four conditions of the Experiment: SR/consistent-first = 5/16, SR/inconsistent-first = 4/14; GR/consistent-first = 5/16; GR/inconsistent-first = 6/16.

Each participant who completed the current IRAP produced 24 response latencies for each test block. Due to a technical fault in the program's recording of the first trial in each block the latency scores for the first trial were removed from all analyses (note, this version of the program was never distributed for general use). The remaining 23 latencies in each test block were converted into the *D*-IRAP scores. For participants who maintained the accuracy and latency criteria across all three pairs of test blocks, the *D*-IRAP scores were calculated as follows:

1) If 10% of a participant's response latencies were less than 300 ms all of the data were removed (no participant had their data removed on this basis);
2) All latencies over 10,000 ms were removed;
3) Twelve standard deviations for the response latencies, calculated for each trial-type, were obtained across the three pairs of test blocks (i.e., blocks 1 & 2; blocks 3 & 4; and blocks 5 & 6);
4) Twenty four mean latencies were calculated, one for each trial-type in each block.
5) A difference score was calculated for each trial-type, in each test block pair, by subtracting the mean latency in the consistent block from the mean latency in the inconsistent block, thus producing 12 difference scores.
6) The difference score for each trial-type in each test block pair was divided by the standard deviation for that trial-type from the corresponding test blocks, resulting in 12 *D*-IRAP scores – one for each trial-type in each pair of test blocks.
7) Four *D*-IRAP scores were calculated, one for each trial-type, by averaging scores across the three pairs of test blocks.

The same general method for calculating *D*-IRAP scores was also applied to the data from participants who had data from a pair of test blocks removed except the algorithm was adjusted accordingly (e.g., 8 standard deviations were calculated in step 3 and 16 mean latencies were calculated in step 4).

The foregoing calculations yielded four mean *D*-IRAP scores for each participant, one for each trial-type. A preliminary 2x2x4 mixed repeated measures analysis of variance (ANOVA) was conducted to determine if the sequence in which the IRAP blocks were presented (i.e., consistent-first versus inconsistent-first) impacted significantly upon the *D*-IRAP effects across the four trial-types and two instruction conditions. The main effect for block order and its interactions with the other two variables were all non-significant (*ps* > .06), and thus this procedural variable was removed from all subsequent analyses.

The overall mean *D*-IRAP scores calculated across participants are presented in Figure 3, divided according to the type of instructions they received (SR versus GR). All *D*-IRAP effects were positive, which indicates that both groups responded more quickly during history-consistent than history-inconsistent blocks for each of the trial-types. In effect, participants tended to respond “True” more quickly than “False” when presented with the label “Color” and the name of a color, and when presented with the label “Shape” and the name of a shape; conversely, participants tended to respond “False” more quickly than “True” when presented with the label “Color” and the name of a shape, and when presented with the label “Shape” and the name of a color.

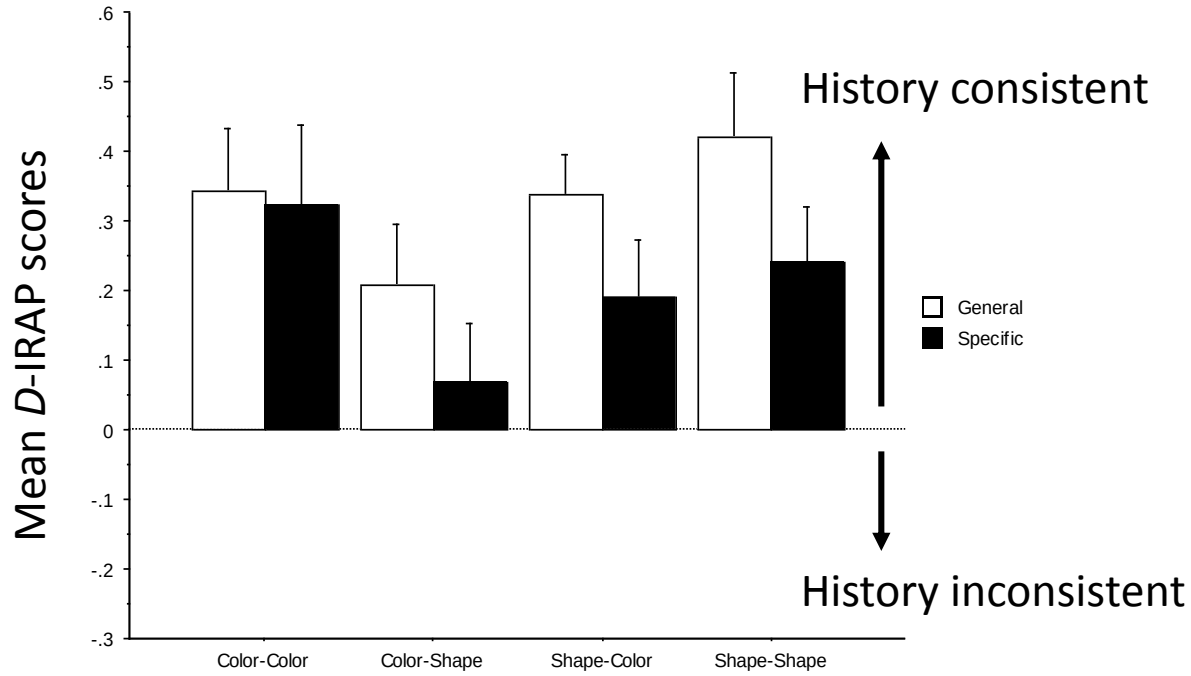

Figure 3. Mean *D*-Scores, with standard error bars, for each trial-type for each condition in Experiment 1. Positive score indicates effects in a history consistent direction.

In comparing the mean *D*-IRAP scores between the SR and GR conditions the *Color-Color* trial-type yielded very similar effect sizes, whereas the remaining three trial-types produced larger effects in the GR relative to the SR condition. In general, the differences between the *Color-Color* trial-type and the other three trial-types were less pronounced for the GR relative to the SR condition. The *D*-IRAP scores were subjected to a 2(*rule-type*) X 4(*trial-type*) mixed repeated measures ANOVA, which failed to yield any significant main or interaction effects (*ps* > .06). Although the effect for rule-type was non-significant, the alpha value was relatively low with an effect size that fell between low and moderate (Cohen, 1988), $F(1,40) = 2.79$, $p = .10$, partial eta squared = .07.

Four one-sample *t*-tests indicated that the *D*-IRAP scores for the GR group each differed significantly from zero; *Color-Color*, $t(20) = 3.85$, $p = .001$; *Color-Shape*, $t(20) = 2.43$, $p = .025$; *Shape-Color*, $t(20) = 5.88$, $p < .0001$; *Shape-Shape*, $t(20) = 4.49$, $p = .0002$.

An additional four one-sample *t*-tests (for the SR-group) yielded significant effects for three of the four trial-types; *Color-Color*, $t(20) = 2.77$, $p < .05$; *Shape-Color*, $t(20) = 2.29$, $p < .05$; and *Shape-Shape*, $t(20) = 3.04$, $p < .01$ (*Color-Shape* $p > .42$).

Overall, therefore, the descriptive statistics suggested that the GR condition yielded stronger *D*-IRAP effects that differed less among the trial-types than was the case for the SR condition. The inferential statistics did not provide firm support for this description of the results, although the one-sample *t*-tests were suggestive (i.e., four significant effects for the GR condition versus only three for the SR condition).

As noted above, only one of the three trial-types produced *D*-IRAP effects that were highly similar across the two rule conditions (the *Color-Color* trial-type). In the introduction it was speculated that providing detailed instructions may serve to produce "genuine" BIRRing on perhaps one or two trial-types, but more EERR-like responses on the remaining trial-types. Insofar as this was the case for the *Color-Color* trial-type for the SR condition in the current experiment, then it may be informative to conduct an inferential statistical analysis with this particular trial-type removed from the data set. In other words, the analysis would be conducted on the three trial-types that may reflect functionally distinct response patterns (i.e., BIRRs versus EERRs). A 2x3 mixed repeated measures ANOVA was therefore conducted, which examined the differences between the two rule conditions across each of the three trial-types (*Color-Shape*, *Shape-Color*, and *Shape-Shape*), and this yielded a significant main effect for rule-condition, $F(1,38) = 4.22$, $p = .046$, partial eta squared = .10. In effect, providing general rules before each block appeared to produce a significantly stronger IRAP effect than when detailed rules were provided.

Of course, this latter set of analyses are rather post-hoc and could be seen as "cherry picking" the data that supports the current theoretical arguments. While recognising that there may well be some substance to the criticism, it is also important to acknowledge a trend in the

data that supports the conclusion that the two rule conditions impacted somewhat differently on the IRAP effects observed in this first experiment. It is also worth bearing in mind that the experiment also suffered from a number of technical and procedural problems and thus the clarity of the effects obtained may have been undermined somewhat by these factors. In view of these various issues, it was decided to replicate the current research while correcting for the problems that it was felt might have contaminated the *D*-IRAP effects obtained therein.

Experiment 2

As noted previously, Experiment 1 suffered from a technical problem, which required that the response latency for the first trial in each block be removed from the data analyses. The impact of losing these data points remains unknown but it is possible that doing so impacted, if only slightly, on the pattern of results obtained. Replicating the current study with an IRAP program that did not contain the "bug" thus seemed wise. In addition, there were two procedural issues associated with Experiment 1 that required attention. First, Experiment 1 was conducted in a group setting with approximately 30 participants completing the IRAP in a large computer laboratory. The standard practice for running IRAP studies, at least at the Maynooth site, is for individual participants to complete the procedure in small sound-attenuated booths on a one-to-one basis with the researcher. Experiment 2 thus adopted this "standard" practice of running participants individually. Second, the two types of instructions provided to participants in the SR condition confounded two variables. That is, the consistent rule was "Colors are colors and shapes are shapes", whereas the inconsistent rule was "Shapes are colors and colors are shapes". In effect, the first word ("Colors" versus "Shapes"), and whether the rule was deemed consistent or inconsistent with common verbal practices, were manipulated across the two rules. Typically, when specific rules have been presented in IRAP studies, only the latter variable has been manipulated. Recent research indicates when specific rules are presented at the beginning of each block in an IRAP, they may have a significant

impact on the resulting IRAP effects (O'Shea, Watson, & Brown, 2015). Thus, in the next Experiment the two specific rules always involved presenting the same word ("Colors") at the beginning.

Method

Participants

67 undergraduate students of psychology at Maynooth University voluntarily participated in the experiment. Participants were not offered any form of remuneration. All participants completed an IRAP and a series of questionnaires. The sample comprised of 33 females and 34 males with an age range of 17-45 yrs. Thirty-nine participants were exposed to the GR condition and 28 were exposed to the SR condition. Eighteen participants failed to complete the IRAP successfully (according to the performance criteria employed in Experiment 1), and thus their data are not presented here. Once again, the attrition rates did not differ dramatically across conditions: SR/consistent-first = 3/14; SR/inconsistent-first = 3/14; GR/consistent-first = 4/18; GR/inconsistent-first = 8/21.

Materials

The IRAP program was similar to the version employed in Experiment 1, except the "bug" that contaminated the first trial in a block was not present and the specific rule for the inconsistent blocks read "Colors are shapes and shapes are colors". The label and target stimuli that were employed in Experiment 1 were also employed in Experiment 2. The demographics questionnaire and rating scales (for shapes and colors) employed in Experiment 1 were also employed in Experiment 2.

Procedure

The procedure was similar to Experiment 1, except that each participant attended the laboratory and completed the IRAP, and the questionnaires, individually in a small sound-attenuated cubicle. The introduction to the IRAP also differed, given that it was not presented

in a group setting. At the beginning of the experiment for each participant the researcher initiated the IRAP program so that it presented the first rule. The researcher then presented the participant with a print out of the four-trial-types from the IRAP, and explained that each one was representative of the tasks that would appear on screen during the experiment. Participants were told that they would respond to each trial-type by pressing either the "D" key for the response option on the left, or the "K" key for the response option on the right. The researcher described the pattern of responding that was required based on the instruction that was presented to the participant. For example, if the instruction read "Colors are colors and Shapes are shapes" the researcher selected a representative trial on the print out – for example "Color" appeared as a label with the word "Red" as a target -- and stated that in this case responding "True" rather than "False" was required. The researcher then described what would happen if participants responded correctly (i.e., the screen would clear and the next trial would be presented) and what would happen if participants responded incorrectly (a red X would appear and a correct response would be required before the program continued to the next trial). When the researcher had worked through the four examples of each trial-type, he invited the participants to press the spacebar and engage with the task in accordance with the instructions.

All procedural features of the IRAP were the same as those described for Experiment 1. Upon completion of the IRAP, all participants completed the questionnaire and rating scales, and thereafter were thanked for their time, debriefed, and dismissed.

Results and Discussion

The data from the IRAP were prepared for analysis in a similar manner to that employed for Experiment 1, except the data point from the first trial in each block was not excluded (because there was no bug in the program). A preliminary 2x2x4 mixed repeated measures analysis of variance (ANOVA) was conducted to determine if the sequence in

which the IRAP blocks were presented (i.e., consistent-first versus inconsistent-first) impacted significantly upon the *D*-IRAP effects across the four trial-types and two instruction conditions. Unlike Experiment 1, block order interacted significantly with the other two main variables, and thus it was not removed from subsequent analyses. The mean *D*-IRAP scores for each trial-type divided according to instruction-type and block-order are presented in Figure 4.

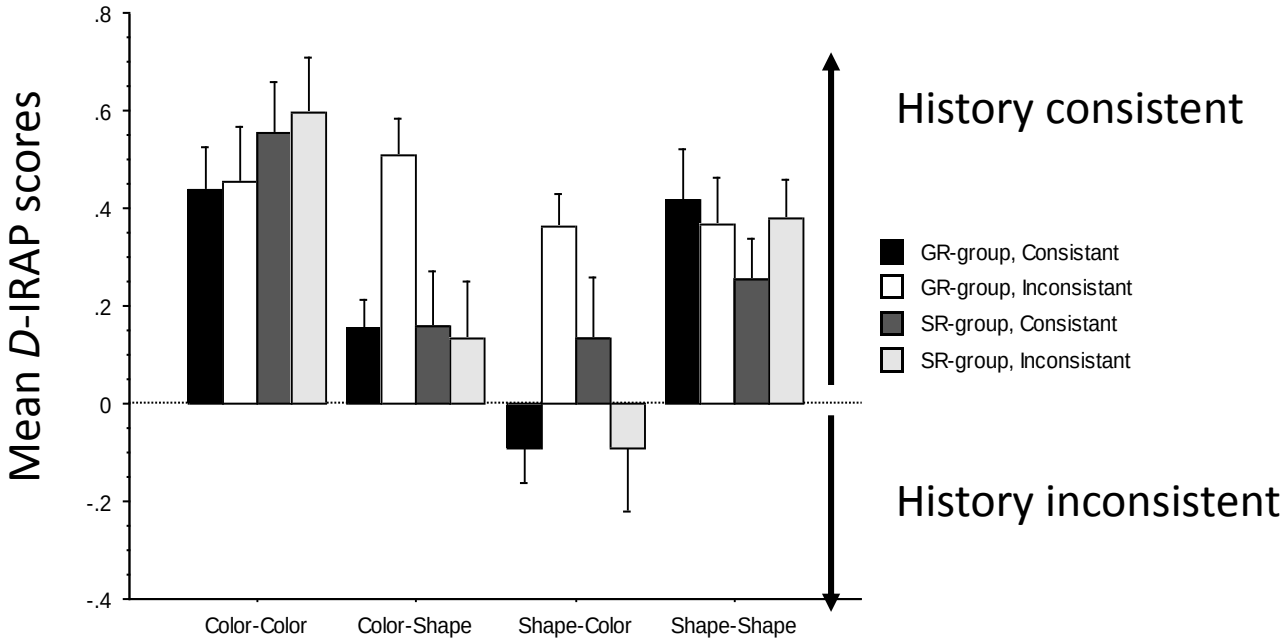

Figure 4. Mean *D*-Scores, with standard error bars, for each group and block order (consistent-first versus inconsistent-first) in Experiment 2. Positive scores indicate effects in a history consistent direction.

The *D*-IRAP effects for the *Color-Color* and *Shape-Shape* trial-types were all relatively strong and in a history-consistent direction; the effects for the *Color-Shape* trial-type were somewhat weaker except for the group who received general instructions and commenced the IRAP with an inconsistent block. The *D*-IRAP effects for the *Shape-Color* trial-type again produced relatively weak effects except for the GR/inconsistent-first group,

but on this occasion two of the weak effects were in a history-inconsistent direction (for the GR/consistent-first and SR/inconsistent-first conditions). Overall, therefore, it appears that the type of instructions and the order in which the blocks were presented impacted upon the *D*-IRAP effects recorded across the four trial-types with the exception of the GR/inconsistent-first group.

The results of the 2x2x4 ANOVA revealed a main effect for trial-type, $F(3,45) = 17.47$, $p < .0001$, partial eta squared = .18, but as noted above this was moderated by a three-way interaction, $F(3,45) = 4.87$, $p = .003$, partial eta squared = .08. The nature of this interaction was explored using four follow-up one-way between-participant ANOVAs for each trial-type and four one–way within-participant ANOVAs comparing the *D*-IRAP scores for the four combinations of instruction-type and block-order. Two of four between-participant ANOVAs proved to be significant, one for the *Color-Shape* trial-type, $F(3,45) = 4.2$, $p = .01$, eta squared = .22, and the other for the *Shape-Color* trial-type, $F(3,45) = 5.07$, $p = .004$, eta squared = .25, but the other two did not ($ps > .6$). Three of the four within-participant one-way ANOVAs each proved to be significant; GR/consistent-first, $F(3, 39) = 9.73$, $p < .0001$, partial eta squared = .43; SR/consistent-first, $F(3, 30) = 5.59$, $p = .004$, partial eta squared = .33; SR/inconsistent-first, $F(3, 30) = 10.31$, $p = .0001$, partial eta squared = .51. The ANOVA for the GR/inconsistent-first condition yielded a non-significant effect with a small effect size, $F(3, 36) = .574$, $p = .63$, partial eta squared = .004. The results of 16 one-sample *t*-tests for each of the *D*-IRAP effects are presented in Table 1. The only condition that produced significant effects across all four trial-types was the GR/inconsistent-first condition; the other three conditions all yielded significant effects for the *Color-Color* and *Shape-Shape* trial-types. Overall, therefore, the inferential statistics confirmed the conclusion, arising from visual inspection of Figure 4, that instructions and block order impacted upon two of the four IRAP trial-types (*Color-Shape* and *Shape-Color*) for three of the four groups.

Table 1.

The t-scores and p-values from the one-sample t-tests for each trial-type effect for each group in Experiment 2

Subject Group	Color-Color	Color-Shape	Shape-Color	Shape-Shape
GR-Con	5.04 (p = .0002)	2.72 (p =.02)	-1.25 (p = .23)	3.96 (p = .002)
GR-Incon	4.04 (p = .001)	6.58 (p = .0001)	5.33 (p = .0002)	3.86 (p = .002)
SR-Con	5.33 (p = .0003)	1.41 (p = .18)	1.08 (p = .31)	3.09 (p = .012)
SR-Incon	5.34 (p = .0003)	1.16 (p = .27)	-.68 (p = .51)	4.74 (p = .0008)

At this point in the research programme it became clear that the type of instructions provided to participants when they complete an IRAP may impact quite dramatically on their performances, but this instructional effect was moderated by the order in which the IRAP blocks were presented. The fact that the specific rule conditions both produced relatively strong *D*-IRAP effects on the *Color-Color* and *Shape-Shape* trial-types is consistent with the argument that the block rules facilitated BIRR-like properties for these two relational response classes. Interestingly, the general rule appeared to produce significant history-consistent *D*-IRAP effects across the four trial-types but only when participants commenced the IRAP with a history-inconsistent block of trials. A possible explanation for this somewhat unexpected outcome will be considered below.

Experiment 3

In the previous two experiments, two types of rules were employed, specific and general. The specific rules clearly focused on particular parts of the relational network and two of the trial-types. Although the general rules did not explicitly specify particular parts of the network (or trial-types) it is possible that the instruction to “respond correctly to the

stimuli", for example, served to highlight the coherent relations within the network and the trial-types. In effect, the word "correct" coordinated with the pre-experimentally established coherent relations (i.e., *Color-Color* and *Shape-Shape*). Consequently, participants may have generated their own "specific" rules, which functioned similarly to the two rules presented in the SR condition. The generation of such rules would likely be enhanced when participants commenced with a block of history-consistent trials because the IRAP would require a pattern of responding that cohered with the participants' pre-experimental verbal histories. The foregoing outcome may have been far less likely when the IRAP commenced with the instruction to "respond incorrectly to the stimuli" (and a block of history-inconsistent trials). Of course, it would be possible for participants to self-generate a rule that cohered with that initial IRAP block (similar to that presented within the SR condition -- "Colors are shapes and shapes are colors"). However, the generation of such a rule may be less likely because it fails to cohere with the participants' pre-experimental histories. More informally, participants may have found it easier to generate a rule that cohered with their natural verbal relations than a rule that did not. Thus, a similar pattern of trial-type effects was observed in the GR and SR conditions (i.e., a mix of significant and non-significant effects), but only when the IRAP commenced with a history-consistent general rule (i.e. "respond correctly to the stimuli").

The foregoing explanation is of course quite speculative, but it does suggest that it should be possible to provide instructions that neither specify parts of the network being assessed, nor cue self-generated rules that serve a similar function. The aim of such instructions would be to generate four significant trial-type effects even when commencing with a history-consistent block of trials. In an effort to achieve this objective, instructions were devised that served to emphasize each trial-type equally. That is, participants were instructed to respond as if "True" is consistent and "False" is inconsistent on history-consistent blocks, and to respond as if "True" is inconsistent and "False" is consistent on

history-inconsistent blocks. These rules do not therefore specify correct or incorrect responding in terms of the network, which may highlight coherent over incoherent relations, but simply told participants how to respond to the response options on each trial-type. Would these rules produce four significant trial-type effects on an IRAP that commenced with a history-consistent block of trials?

Participants

Seventeen undergraduate students at Maynooth University voluntarily participated in the experiment. Participants were not offered any form of remuneration. All participants completed an IRAP, the questionnaire, and the rating scales. The sample comprised of 11 females and 6 males with an age range of 20-30 yrs. Five participants failed to complete the IRAP successfully and their data are not reported here.

Materials

The IRAP program was similar to the version employed in Experiment 2, except the rules provided to the participants focused on the functions of the response options rather than the functions of the labels and targets for a particular block (see below). The label and target stimuli that were employed in Experiments 1 and 2 were also employed in Experiment 3. The demographics questionnaire and rating scales (for shapes and colors) employed in Experiments 1 and 2 were also employed in Experiment 3.

Procedure

All procedural features of the IRAP were the same as those described for Experiment 2, except that all participants commenced with a history-consistent blocks of trials. In addition, the instructions were modified in order to focus on each trial-type equally by emphasizing the function of the response options. That is, before each history-consistent block of trials the instruction read “Please respond as if true is consistent and false is inconsistent”; before each history-inconsistent block of trials the instruction read “Please

respond as if true is inconsistent and false is consistent." Upon completion of the IRAP, all participants completed the questionnaires, and thereafter were thanked for their time, debriefed, and dismissed.

Results and Discussion

The data from the IRAP were prepared for analysis in the same manner to that employed for Experiment 2. The mean *D*-IRAP scores for each trial-type are presented in Figure 5, which shows that all four effects were relatively strong and in a history-consistent direction; the effects for the *Color-Color* and *Shape-Color* trial-types were stronger than for the other two.

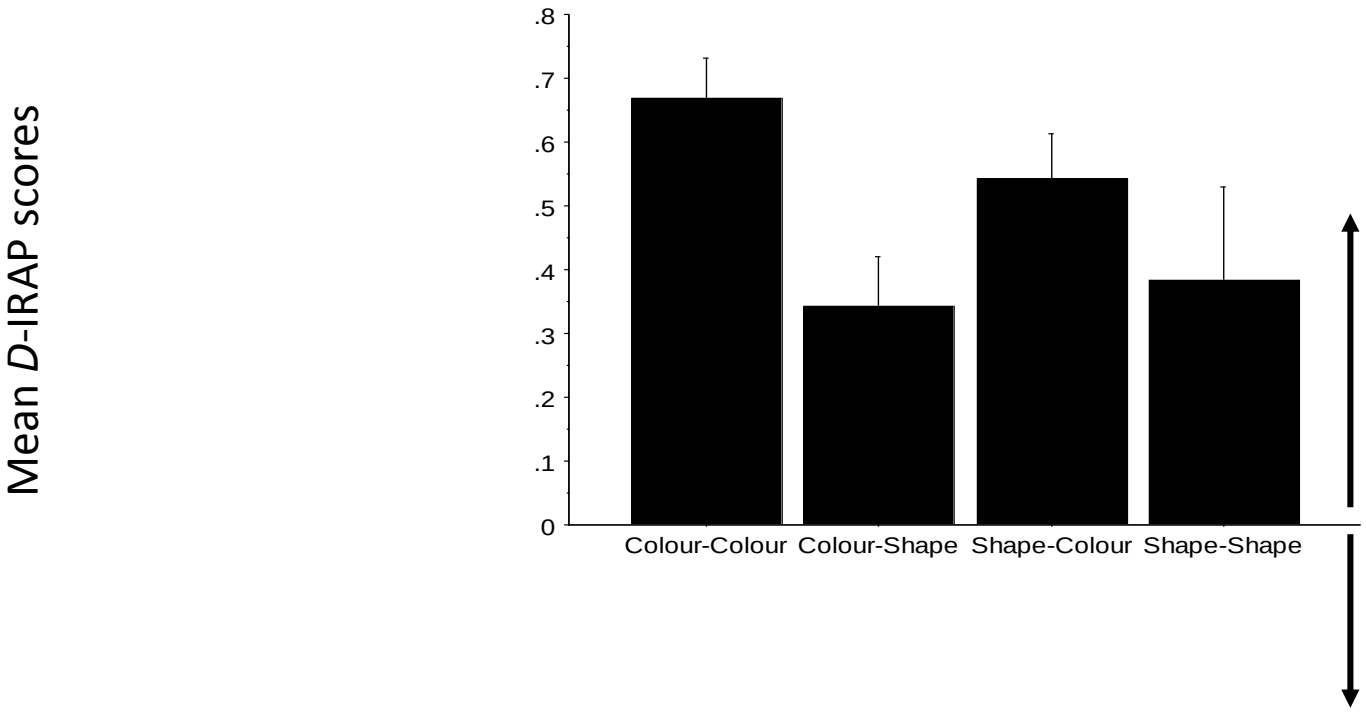

Figure 5. Mean D-Scores, with standard error bars, for each trial-type in Experiment 3. Positive scores indicate effects in a history consistent direction.

The results of a one-way within-group ANOVA revealed a marginally significant effect for trial-type, $F(3,33) = 2.81$, $p = .055$, partial eta squared = .07. Differences between trial-types were investigated by means of paired *t*-tests and these revealed that the *D*-IRAP

effects for the *Color-Color* trial-type differed significantly from the *Color-Shape, t*(11) = 4.14, p < .01, and *Shape-Shape*, *t*(11) = 2.18, *p* = .05, trial-types (remaining *ps* > .14). Four one-sample *t*-tests confirmed that each of the trial-type effects differed significantly from zero; *Color-Color t*(11) = 10.52, *p* <.0001; *Color-Shape t*(11) = 4.24, *p* = .001; *Shape-Color t*(11) = 7.66, *p* < .0001; *Shape-Shape t*(11) = 2.65, *p* = .02. In summary, therefore, providing response-focused instructions appeared to generate relatively strong (and significant) IRAP effects for each of the four trial-types, when participants commenced the procedure with a history-consistent block of trials.

General Discussion

Overall, the current study indicates that providing detailed rules that specify particular parts of a relational network, and trial-types within the IRAP, may have quite dramatic effects on IRAP performances. Furthermore, the findings suggested that even when general rules are presented, a participant's initial contact with the IRAP contingencies (history-consistent versus inconsistent) may moderate the IRAP effects quite dramatically. When an interpretation of this moderating effect (i.e., participants produced self-directed rules) was tested with a third and final experiment, this provided evidence to support our speculative explanation.

At the end of the introduction, it was suggested that presenting a rule or instruction that specified only particular parts of a relational network may generate IRAP effects that involve relational responding that is less BIRR-like than would be suggested by the REC model. The pattern of IRAP effects observed across the three studies reported here provide some support for this argument. At this point, it seems important to provide a potential explanation for the full pattern of effects observed in the current research. The main finding that appears to require a clear explanation is the fact that *Color-Color* and *Shape-Shape* trial-types tended to produce relatively strong, history-consistent effects across all conditions, but

the *Color-Shape* and *Shape-Color* trial-types did not. In effect, providing a detailed rule did not appear to undermine the IRAP effect for all four trial-types, but only for one or two of them. Why might this be the case?

One explanation is that given a history-consistent rule the IRAP involves relating a "same" relation to another "same" relation on trial-types 1 and 4, but relating a "same" relation with a "difference" relation on trial-types 2 and 3. That is, given the rule, "colors are colors and shapes are shapes" the first part of the rule coheres with trial-type 1 and the second part of the rule coheres with trial-type 4. Insofar as relational coherence is frequently reinforced within the wider verbal community, relational responding on these trial-types may occur relatively rapidly (i.e., it is BIRR-like). In contrast, neither parts of the rule cohere with trial-types 2 and 3 (i.e., *Color-Shape* and *Shape-Color*), which may serve to undermine BIRR-like responding. Now consider the history-inconsistent blocks. Trial-types 1 and 4 do not cohere with the rule, which may well reduce BIRR-like responding, relative to history-consistent blocks. Consequently, trial-types 1 and 4 will tend to yield relatively strong trial-types effects in a history-consistent direction. For trial-types 2 and 3, however, although they cohere with the rule for responding on the IRAP they do not cohere with the relational networks established by the wider verbal community (e.g., colors are not in fact shapes). Consequently, the BIRR-like properties of the relational responses produced on these two trial-types may also be undermined during the history-inconsistent blocks of trials. In summary, therefore, the specific rules presented in the current studies may have supported BIRR-like responding for trial-types 1 and 4 during history-consistent blocks but not during history-inconsistent blocks; in contrast, BIRR-like responding may have been much reduced or entirely absent for trial-types 2 and 3 during both history-consistent and history-inconsistent blocks. Insofar as this was the case one would expect predictable IRAP effects

for trial-types 1 and 4 but unpredictable effects for trial-types 2 and 3, which is what was observed.

In offering the foregoing conceptual explanation for the observed pattern of results it must be recognized that it is rather speculative and important questions remain. For example, one finding that raises questions is the fact that although all four IRAP effect in Experiment 3 were significantly different from zero they did differ significantly from each other. Given that participants were provided with response-focused instructions, rather than rules which specified particular parts of a relational network, why did these differences emerge? One possibility is that even under these instructional conditions some participants' self-generated rules that were broadly similar to those presented during the SR conditions in Experiments 1 and 2. In making this argument it is important to bear in mind that many of the participants who completed Experiment 3 had previously participated in at least one IRAP study, which involved the provision of detailed rules. It is possible, therefore, that this pre-experimental history with the IRAP, and the provision of rules specifically, encouraged some participants to generate specific rules during Experiment 3, thus generating a somewhat uneven pattern of IRAP effects. Perhaps future research could explore the impact of previous IRAP exposures with and without detailed rules on subsequent IRAP performances.

The findings obtained across the three studies raise a number of important questions for the use of the IRAP as a measure of so called implicit cognition, and as a tool for measuring relational framing "in-flight". Clearly, the types of rules that are presented to participants during an IRAP are not an insignificant variable that have little or no impact on performance. The instructions or rules appear to be quite impactful and indeed interact with other IRAP variables, such as the order in which the blocks are presented (history-consistent versus history-inconsistent) and possibly with prior exposures to other IRAPs. Of course, it is worth noting at this stage that the current research did not attempt to assess the impact of rules

or instructions on the predictive validity of the IRAP effects (see Vahey, Nicholson, & Barnes-Holmes, 2015). For example, it has yet to be determined if strong IRAP effects have greater predicative validity than weaker IRAP effects, or indeed vice versa. Nevertheless, the fact that the current research has shown that rules do impact on IRAP performances highlights that future research will need to explore this area systematically if the IRAP, and its derivatives (e.g., De Houwer, Heider, Spruyt, Roets & Hughes, 2015; Levin, Hayes, & Waltz, 2010; O'Shea, et al., 2015), and indeed perhaps other implicit measures, are to be used with increasing precision and confidence over the coming years.

In pursing the current line of research a number of issues raised by the current experiments seem particularly important. First, the rates of attrition due to failure to reach the specific performance criteria for the IRAP across the current experiments was particularly high relative to many other published IRAP studies (see Hughes & Barnes-Holmes, 2013). The reason for the high attrition remains unclear at the present time, but the most likely explanation is that there was no opportunity for the researcher to offer additional instructions for clarification to participants at the beginning of the IRAP because the instructions were the very focus of the research.

Second, a related issue pertains to how the introductory rules or instructions employed in the current research compare to those used in previous IRAP studies. The use of quite specific rules presented before each block of trials was only introduced with the 2012 version of the IRAP program and thus it is a relatively recent modification. On balance, it is difficult if not impossible to know exactly what informal rules and prompting were employed in earlier IRAP studies and thus a relatively precise post-hoc comparison is not realistic. But given the current findings, a systematic experimental analysis now certainly seems warranted.

Third, the key rationale of the current study was to examine the impact of different types of introductory rules on IRAP performances. Conceptually, we have speculated that the

observed rule effects emerge, at least in part, because they impact upon the amount and type of relational activity that occurs on each IRAP trial. On balance, no measures were employed in the current experiments that aimed to assess this activity directly. Simply relying on a latency-based metric is limited in the sense that even if a procedure requires a response within a certain temporal window, as is the case with the IRAP, more or less relational activity may occur within that window without necessarily impacting on latency per se. By way of analogy, two different runners could cover more or less distance in the *same amount of time,* and thus time *per se* would not reflect number of strides taken by each runner. One possible direction for future research, therefore, would be to employ a "think aloud" procedure and protocol analyses, summarized by Hayes (1986), which has since been employed by behavior-analytic researchers studying human behavior in other contexts (e.g., Wulfert, Dougher, & Greenway, 1991; Cabello, Luciano, Gomez, & Barnes-Holmes, 2011). Insofar as using such a methodology would provide direct access to participants' private relational activity as they encountered each IRAP trial, we could learn a great deal about the way in which different rules impact upon the dynamic interactions that occur between the IRAP and participants' behavior.

Compliance with Ethical Standards

Conflict of Interest: Martin Finn declares that he has no conflict of interest. Dermot Barnes-Holmes declares that he has no conflict of interest. Ian Hussey declares that he has no conflict of interest. Joseph Graddy declares that he has no conflict of interest.

Ethical Approval: All procedures performed in studies involving human participants 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.

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

References

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(3), 527-542.

Barnes-Holmes, D., Barnes-Holmes, Y., Hussey, I., & Luciano, C. (in press). Relational frame theory: Reflecting upon its historical roots and future development. In R. Zettle, S. C. Hayes, D. Barnes-Holmes, & T. Biglan (Eds), *Wiley Handbook of contextual behavioural science.* Cambridge: Wiley-Blackwell.

Barnes-Holmes, D., Hayes, S.C., Dymond, S., & O'Hora, D. (2001). Multiple stimulus relations and the transformation of stimulus functions. In S.C. Hayes, D. Barnes-Holmes, & B. Roche (Eds.), *Relational Frame Theory: A post-Skinnerian Theory of Language and Cognition* (pp. 51-72). New York: Kluwer Academic.

Cabello, F., Luciano, C., Gomez, I., & Barnes-Holmes, D. (2011). Human schedule performance, protocol analysis, and the" silent dog" methodology. *The Psychological Record*, *54*(3), 6.

Cohen, J. (1988). Statistical power analysis for the behavioural sciences. New York: Academic Press.

De Houwer, J., Heider, N., Spruyt, A., Roets, A., & Hughes, S. (2015). The relational responding task: toward a new implicit measure of beliefs. *Frontiers in Psychology*, 6, 319.

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

Hayes, S. C. (1986). The case of the silent dog - verbal reports and the analysis of rules: a review of Ericsson and Simon's protocol analysis: verbal reports as data. *Journal of the Experimental Analysis of behavior*, *45*(3), 351-363.

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

Hughes, S., & Barnes-Holmes, D. (2011). On the formation and persistence of implicit attitudes: New evidence from the implicit relational assessment procedure (IRAP). *The Psychological Record*, 61, 391-410.

Hughes, S., & Barnes-Holmes, D. (2013). A functional approach to the study of implicit cognition: The Implicit Relational Assessment Procedure (IRAP) and the Relational Elaboration and Coherence (REC) model. In S. Dymond & B. Roche (Eds.), Advances in relational frame theory: Research and application (pp. 97–125). Oakland CA: New Harbinger.

Hughes, S, & Barnes-Holmes, D. (in press). Relational frame theory: the basic account. In R.

Zettle, S. C. Hayes, D. Barnes-Holmes, & T. Biglan (Eds), *Wiley Handbook of contextual behavioural science.* Cambridge: 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, 17-38.

Hughes, S., Hussey, I., Corrigan, B., Jolie, K., Murphy, C., & Barnes-Holmes, D., (2016). Faking revisited: Exerting strategic control over performance on the Implicit Relational Assessment Procedure. Manuscript submitted for publication.

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.

Levin, M.E., Hayes, S.C. & Waltz, T. (2010). Creating an implicit measure of cognition more suited to applied research: A test of the Mixed Trial – Implicit Relational Assessment Procedure (MT-IRAP). International Journal of Behavioral Consultation and Therapy, 6, 245-262.

O'Shea, B., Watson, D. G., & Brown, G. D. (2015). Measuring implicit attitudes: a positive framing bias flaw in the Implicit Relational Assessment Procedure (IRAP). *Psychological Assessment*. Advance online publication. http://dx.doi.org/10.1037/pas0000172

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.

Wulfert, E., Dougher, M. J., & Greenway, D. E. (1991). Protocol analysis of the correspondence of verbal behavior and equivalence class formation. *Journal of the Experimental Analysis of Behavior*, *56*(3), 489-504.